Whole genome RNA expression profiling for the identification of novel biomarkers in the diagnosis and prognosis of biliary tract cancer

Dr Michael Huw Chapman

Thesis submitted for the award of PhD

Primary supervisor
Dr Stephen Pereira
Secondary supervisor
Dr James Dooley
Laboratory mentor
Dr Fausto Andreola

Institute of Hepatology, Faculty of Biomedical Sciences,
University College London.

March 2011

Declaration

I, Michael Huw Chapman, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

The studies described in this thesis were made possible by a number of individuals to whom I am indebted. In particular, Dr Stephen Pereira (primary supervisor), Dr James Dooley (secondary supervisor) and Dr Fausto Andreola (laboratory mentor) were available for advice and support.

Others who have greatly assisted are Dr Neomal Sandanayake for aiding with our sample collection and storage, Dr George Webster and Dr Adrian Hatfield for help with obtaining samples at ERCP and Dr Virginie Cerec for support with cell culture and other laboratory work. Dr Dipok Dhar assisted with data analysis and statistical support. Histopathology assessments of tissue sections were kindly reported by Dr Maesha Deheragoda. All are staff at UCL or University College London Hospitals NHS Foundation Trust.

Robert Tidswell helped carry out some of the experiments in this thesis as part of an intercalated BSc in Clinical Sciences at UCL (2008/2009). His contribution and assistance is appreciated.

The studies were supported by a grant from the British Liver Trust and research and discretionary funds of Dr Stephen Pereira and Dr James Dooley (both UCL Medical School). I gratefully acknowledge this support.

Abstract

Biliary tract cancer (BTC) is difficult to diagnose, in part related to the lack of reliable tumour markers. The aim of this project was to use whole genome RNA expression profiling in order to identify novel biomarkers for diagnosis and prognosis in biliary tract cancer.

Chapter 1 summarises clinical aspects of BTC as well as current diagnostic and prognostic tests.

Chapter 2 addresses the identification of circulating tumour cells for the diagnosis of BTC. It includes details of a study investigating measurement of circulating cytokeratin 19 fragments (CYFRA 21-1), demonstrating that CYFRA 21-1 is a more specific, but less sensitive diagnostic marker than CA19-9, and predicts a poor prgnosis in BTC.

Chapter 3 investigates the potential for using RNA isolated from archived formalin fixed, paraffin embedded (FFPE) surgical and explanted liver tissues from patients with primary sclerosing cholangitis (PSC) with and without cholangiocarcinoma, for use in whole genome RNA expression analysis. We demonstrate that, although technically possible, the rarity of samples and RNA degradation that occurs as a result of the tissue processing, are such that further evaluation using these materials is not feasible at this time.

Chapter 4 addresses and validates methodology for isolating RNA from samples of biliary brushings taken at the time of endoscopic retrograde

cholangiopancreatography (ERCP). We demonstrate that RNA isolated from biliary brushings is of low quantity and degraded, and that this degradation occurs *in vivo*. However, we demonstrate that such RNA is still useful for downstream applications such as quantitative real time PCR and is therefore suitable for whole genome RNA expression analysis using microarray technology.

Chapter 5 describes the methods and results obtained from using whole genome RNA expression analysis using microarray of RNA isolated from ERCP biliary brushings. The results are presented as a shortlist of candidate genes requiring further validation.

Chapter 6 provides results of qPCR studies performed in order to validate the gene expression profile identified by microarray. A selection of candidate genes are investigated using TaqMan Array and SYBR Green qPCR and demonstrate a high correlation with the pattern of expression shown by microarray.

Chapter 7 investigates whether a selection of the genes identified in malignant biliary brushings are similarly upregulated in fresh frozen surgical resection material from patients with benign and malignant biliary diseases. In addition, we provide evidence for gene translation and upregulation at the protein level by immunohistochemistry for a selection of the protein products.

Chapter 8 discusses the main conclusions drawn from the work as well as potential future studies.

Table of Contents

List of figures

List of Tables

List of Abbreviations

AUC	Area under curve
BTC	Biliary tract cancer
CA 19-9	Carbohydrate antigen 19-9
CEA	Carcinoembryonic antigen
CC	Cholangiocarcinoma
cDNA	Complementary DNA
CK19	Cytokeratin 19
CRUK	Cancer Research UK
CT	Computed tomography
Ct	Cycle threshold
DNA	Deoxyribonucleic acid
ERCP	Endoscopic retrograde cholangiopancreatography
FFPE	Formalin fixed, paraffin embedded
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GBCa	Gallbladder cancer
H&E	Haematoxylin and Eosin
IHC	Immunohistochemistry
LCM	Laser capture micro-dissection
MRCP	Magnetic resonance cholangiopancreatography
NPV	Negative predictive value
PCR	Polymerase chain reaction
PPV	Positive predictive value
PSC	Primary sclerosing cholangitis
qPCR	Quantitative real time PCR
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
UCH	University College Hospital
UCL	University College London

Chapter 1.

1 Biliary tract cancer

1.1 Introduction

Biliary tract cancer (BTC) includes the primary liver cancers cholangiocarcinoma and gallbladder cancer, which are both thought to arise from biliary epithelial cells. Cholangiocarcinoma (CC) is usually classified according to its location as intra-hepatic or extra-hepatic (Figure 1) but some older classifications and publications include a separate peri-hilar type (intra-hepatic in current classification) which has caused confusion in estimating incidence rates and understanding the biology of the disease (Welzel, McGlynn et al. 2006; Matull, Khan et al. 2007; Blechacz, Sanchez et al. 2009). Peripheral, intra-hepatic CC tends to form mass lesions within the hepatic parenchyma. It is more common in East Asia and has a strong association with chronic infestation with the liver fluke (*Opisthorchis or Clonorchis spp*). Extra-hepatic CC is the commoner type in the western world and tends to infiltrate along the major bile ducts, but may also form mass lesions. Extra-hepatic CC is likely to be a biologically different disease to the peripheral type more commonly seen in East Asia. The incidence of cholangiocarcinoma appears to be rising with an incidence of approx 1000 cases per year in the United Kingdom in 2001 and more recent data from the Office of National Statistics recording 2650 cases in 2005. (Taylor-Robinson, Toledano et al. 2001; Toledano, Mehtan et al. 2009). The rising incidence is partly related to a rise in the intra-hepatic type, for which the cause is unknown but may also be

related to alteration in disease classification or diagnostic accuracy (Taylor-Robinson, Toledano et al. 2001; Khan, Taylor-Robinson et al. 2002).

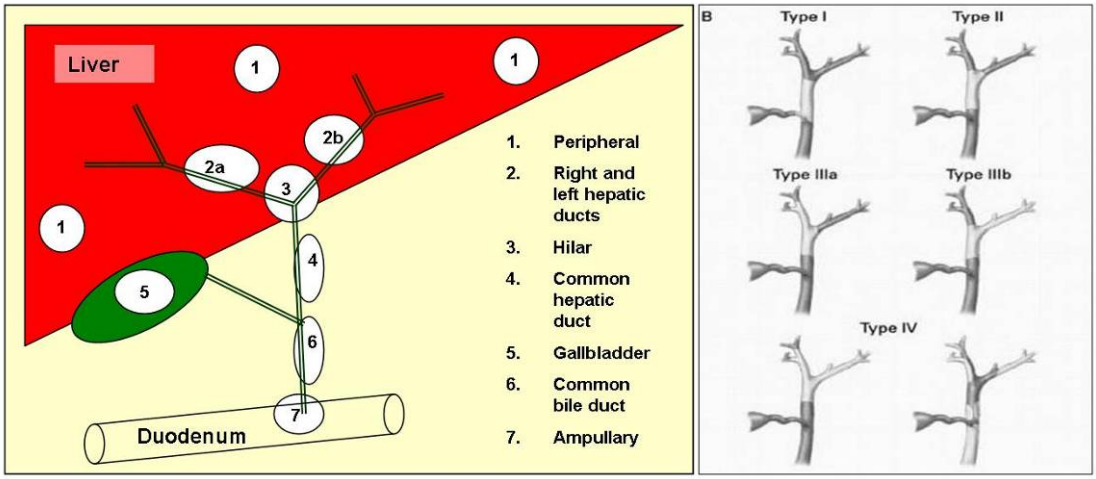


Figure 1. Left: Anatomy of the main hepatobiliary system with description of main anatomical classification of biliary tract cancers. Right: Bismuth classification of extra-hepatic CC (reproduced from (de Groen, Gores et al. 1999).

The pathobiology of BTC is poorly understood. Most cases of extra-hepatic cholangiocarcinoma are sporadic with no known cause. Identifiable risk factors account for a small percentage of cases but include age, smoking, choledochal cysts, exposure to Thorotrast X-ray contrast medium and causes of chronic biliary inflammation including chronic typhoid, *Opisthorchis* infection, choledocholithiasis and primary sclerosing cholangitis which carries a high lifetime risk of up to 20% (Bergquist, Ekbom et al. 2002) and remains the most significant known risk factor for cholangiocarcinoma in Western countries.

Cholangiocarcinoma usually presents with symptoms and signs of biliary obstruction (jaundice, dark urine and pruritis), abdominal pain, weight loss and/or features of cholangitis (fevers, rigors and pain). There are few

symptoms or signs in early disease and as a result, disease usually presents at an advanced stage.

One of the greatest challenges in BTC is the difficulty in diagnosis, particularly differentiating benign from malignant biliary strictures in conditions such as primary sclerosing cholangitis. The growth of CC follows 3 forms; 1) peri-ductal infiltrating, 2) papillary or 3) mass forming (Lim and Park 2004). Spreading intra-ductal disease is more common than the mass forming type so that there is often no mass lesion identified on imaging, adding to the difficulties in diagnosis.

Patients with BTC have a poor prognosis, in part related to difficulties and delays in diagnosis. Most patients present at an advanced stage with five year survival rates of less than 5% and median survival of only six to nine months. The only potentially curative treatment is hepatic resection or, in highly selected cases, liver transplantation (Khan, Davidson et al. 2002; Becker, Rodriguez et al. 2008). Series from the US suggest that highly selected cases in PSC disease have 5 year survival rates of up to 82% when treated with preoperative chemotherapy and radiotherapy followed by transplantation (Rea, Heimbach et al. 2005; Wu, Johlin et al. 2008). However, in most countries (including the UK), CC remains a contraindication to liver transplantation because of the high rates of recurrence observed in older series (Khan, Davidson et al. 2002).

Most treatments for CC are given with palliative intent. Results of radiotherapy have been disappointing and there is currently little role for radiotherapy in the treatment of cholangiocarcinoma in the UK (Khan, Davidson et al. 2002). CC has been considered a relatively chemo-resistant cancer with moderate response rates of up to 30% reported with gemcitabine and 5-fluorouracil (5-FU) based regimen (Thongprasert 2005). However a recent report of the largest chemotherapy study undertaken to date in BTC (ABC-02), shows significant progress with chemotherapy (Valle, Wasan et al. 2010). In this phase III study, 410 patients with locally advanced biliary tract cancer (CC, gallbladder and ampullary cancer) were randomised to treatment with gemcitabine plus cisplatin over gemcitabine alone. The patients treated with combination of gemcitabine plus cisplatin had a significant survival advantage (11.7 months versus 8.2 months, $p=0.002$) and better progression-free survival (8.0 months versus 5.0 months, $p<0.001$) than in the gemcitabine alone arm, without significant differences in adverse events between the two groups (Valle, Wasan et al. 2010). As a result, chemotherapy with the combination of gemcitabine and cisplatin is the current standard of care for suitable patients with BTC. Combination of chemotherapy with other systemic therapies such as small molecule or antibody blockade of tyrosine kinase receptors may further improve treatment regimens in the future.

More novel physical and biological therapies including photodynamic therapy (Chapman and Pereira 2009) and the use of biological agents such as tyrosine kinase inhibitors are currently being assessed in phase II and III

clinical trials including the PHOTOSTENT-02 trial coordinated by the CRUK and UCL Cancer Trial Centre (Dr Pereira, Principal investigator).

Biliary stent insertion remains the most common palliative intervention. It has no impact on tumour progression but aims to improve a patient's quality of life by relief or prevention of jaundice, pruritis and cholangitis. Stents may be placed endoscopically or percutaneously, depending on the nature of the stricture and ease of endoscopic access. Both metal and/or plastic stents may be used once pathological diagnosis has been confirmed. There are few randomised controlled trials comparing outcome using plastic or metal stents in BTC (Wagner, Knyrim et al. 1993), but based on studies in pancreatic cancer, metal stents may provide better palliation, result in fewer procedures and are cost effective in those with a life expectancy over 6 months (Flamm, Mark et al. 2002). Either or both the percutaneous and endoscopic routes may be suitable for insertion of biliary stents, but failure to achieve adequate drainage leads to a poor outcome in those patients. A study of 85 patients undergoing percutaneous or endoscopic metal stent insertion showed that in those with inadequate biliary drainage, the prognosis was much poorer (median 1.8 months versus 8.7 months, p=<0.001)(Paik, Park et al. 2009)

The difficulty in diagnosing BTCs is a particular issue in patients with primary sclerosing cholangitis (PSC). In this condition widespread benign bile duct structuring is characteristic. However, CCA is not uncommon and differentiating benign from malignant biliary strictures is a regular clinical challenge. PSC is an immune-mediated chronic liver disease characterised by

inflammation, fibrosis and destruction of intra- and/or extra-hepatic bile ducts leading to cholestasis, bile duct strictures and hepatic fibrosis, which in turn may progress to cirrhosis, portal hypertension and hepatic decompensation (Chapman, Arborgh et al. 1980). One of the biggest clinical problems in managing patients with PSC is the difficulty in differentiating benign from malignant biliary strictures. The incidence of cholangiocarcinoma in PSC is between 0.6% and 1.5% per year with a prevalence of 6-13% and a lifetime risk of up to 20% (Bergquist, Ekbom et al. 2002; Boberg, Bergquist et al. 2002; Burak, Angulo et al. 2004). In up to half of cases, cholangiocarcinoma is identified within a year of diagnosis of PSC and may be the reason for presentation of previously unrecognised PSC (Chapman, Witmann et al. 2008) (Abtrast 1, Figure 73, Appendix). The incidence of cholangiocarcinoma is highest in those with dominant strictures with up to 76% located within the peri-hilar region (Ahrendt, Pitt et al. 1999). To date, no cholangiocarcinoma's have been reported in patients with small duct PSC except for a single report of a patient progressing to classical large duct PSC prior to developing cholangiocarcinoma (Bjornsson, Olsson et al. 2008). Therefore, most cases of CC should be within reach of ERCP sampling. It is often not possible to distinguish benign from malignant biliary strictures using imaging alone, such as magnetic resonance cholangiopancreatography (MRCP). Biliary brush cytology at ERCP is the standard investigation for possible malignant biliary stricture but despite good specificity, reported sensitivity in most centres is poor. Our experience of CC in PSC are summarised in abstract form in the appendix (Abstract 1, Figure73 (Chapman, Witmann et al. 2008)).

1.2 Current diagnostic investigations for biliary tract cancer

As discussed above, the diagnosis of cholangiocarcinoma is well recognised as being a difficult clinical problem. As with many cancers, diagnosis requires cytologic or histologic confirmation of adenocarcinoma with immunostaining in keeping with biliary tract cancer. However, this involves invasive procedures with significant risk of complications. Other indicators of disease, such as circulating biomarkers or improved imaging techniques, would greatly enhance clinical decision making and improve diagnostic algorithms. Current diagnostic tests for BTC are summarised below.

1.2.1 Imaging

1.2.1.1 Ultrasound

Ultrasound is normally performed as the first imaging modality in patients with abnormal liver biochemistry. Mass lesions are not typical in cholangiocarcinoma and therefore are usually not seen on ultrasound scanning. Dilatation of intra-hepatic bile ducts in the presence of normal calibre extra-hepatic bile ducts is suggestive of biliary stricturing but does not differentiate benign from malignant disease. Bile duct thickening may be seen but is non-specific. Other findings such as the identification of hepatic venous or arterial thrombosis may assist in assessing staging and prognosis.

1.2.1.2 Computed tomography (CT)

CT is more sensitive in identifying mass lesions but has similar limitations to ultrasound when no mass lesion is present. CT is useful for staging disease in order to plan and monitor treatment. Studies assessing the accuracy of CT in determining resectability of CC report sensitivities and specificities of 94-100% and 48-79% respectively (Tillich, Mischinger et al. 1998; Cha, Han et al. 2000; Lee, Kim et al. 2006; Aloia, Charnsangavej et al. 2007). Multi-slice CT cholangiography has improved the diagnostic accuracy for benign and malignant biliary disease. A prospective study (n=36) comparing magnetic resonance cholangiopancreatography (MRCP) with CT cholangiography demonstrated an accurate diagnosis using CT in 34 of the 36 cases of biliary obstruction (benign and malignant pancreatobiliary disease) (Zandrino, Curone et al. 2005). A further study of CT cholangiography in 34 patients with biliary obstruction reported accurate diagnosis of 93% of patients with benign disease and 94% in those with malignant obstruction (pancreatic and biliary cancer) (Ahmetoglu, Kosucu et al. 2004). However, these studies included both pancreatic and biliary malignancy and the experience of most clinicians is that imaging is far less accurate in routine clinical practice, particularly in cases of cholangiocarcinoma.

1.2.1.3 Magnetic resonance imaging (MRI) and magnetic resonance cholangiopancreatography (MRCP)

The non invasive nature of MRI/MRCP means that it is considered the investigation of choice for imaging biliary anatomy. Studies investigating the

accuracy of MRCP in differentiating benign from malignant biliary obstruction have been mostly performed in patients with PSC. The largest study in 230 patients with PSC (23 with CC) reported a sensitivity of 63% and specificity of 79% with a PPV of 40% for the diagnosis of CC (Charatcharoenwitthaya, Enders et al. 2008). In unselected biliary disease, MRCP is reported to have sensitivities of 48-88% and specificities of 71-95% for the diagnosis of malignant biliary strictures (Rosch, Meining et al. 2002; Romagnuolo, Bardou et al. 2003; Domagk, Wessling et al. 2004).

1.2.1.4 Endoscopic retrograde cholangiopancreatography (ERCP)

ERCP is an important investigation for the investigation of suspicious biliary strictures (Figure 2). It provides accurate delineation of the pattern and extent of stricturing, is sensitive in assessment of secondary complications such as the presence of stones, and provides access for tissue sampling of bile and biliary strictures. However, ERCP has a risk of complications and is therefore only indicated when therapeutic interventions such as brush cytology, stricture dilatation or stent insertion are required.

The most difficult clinical scenario with regards investigating strictures is differentiating benign from malignant strictures in patients with PSC, although similar problems arise for many causes of biliary strictures. Studies investigating the accuracy of ERCP over MRCP for the diagnosis of CC in PSC suggest that both have comparable accuracy. These data come from

studies primarily assessing the role of MRCP in diagnosis of benign PSC disease which report sensitivities and specificities of 83%to 100% and 92% to 100% of MRCP for the diagnosis of PSC (Ernst, Asselah et al. 1998; Angulo, Pearce et al. 2000; Textor, Flacke et al. 2002; Berstad, Aabakken et al. 2006). Since cholangiography, either by MR or the endoscopic approach, is insufficient for diagnosis of malignancy, the main role for ERCP in diagnosis remains its ability to obtain tissue samples for diagnosis.

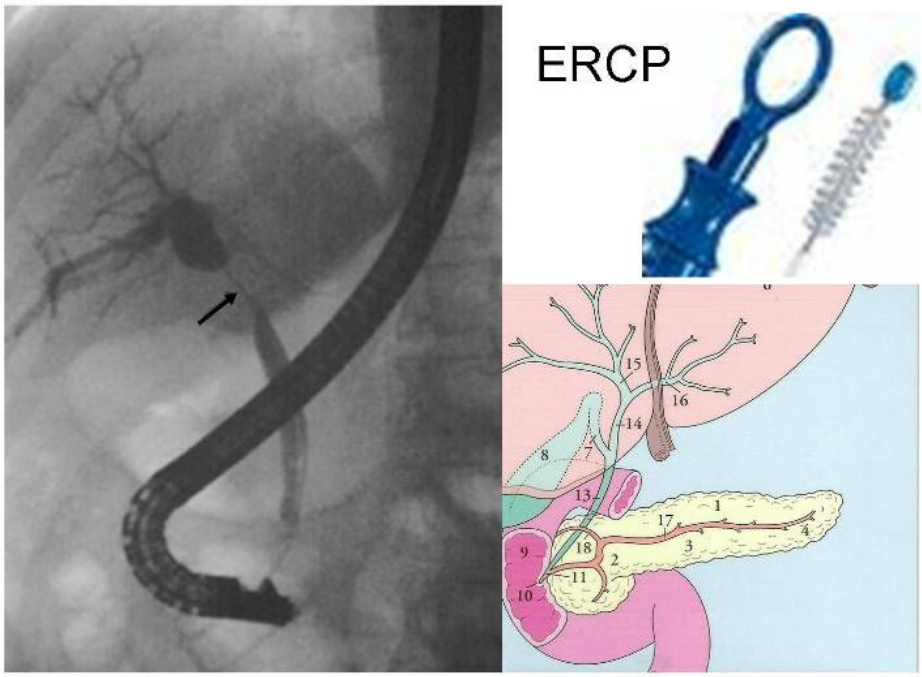


Figure 2 Endoscopic retrograde cholangiopancreatography (ERCP) showing a cholangiogram of a stricture (arrow) in the common hepatic duct (left), endobiliary brush for cytology sampling (top right) and diagram of the local anatomy (bottom right, image from Medical Illusatration Group, Chelsea and Westminster Hospital).

1.2.1.5 Cholangioscopy

Development of ultra-thin fibre-optic fibres has allowed the manufacture of small calibre cholangioscopes introduced through the papilla at ERCP. These

allow direct visualisation of suspicious biliary strictures in order to further characterise and direct biopsies towards abnormal tissues. Early studies investigating the accuracy of cholangioscopy reported a sensitivity of 71-100% and specificity 86-100% for the diagnosis of malignant biliary strictures (Fukuda, Tsuyuguchi et al. 2005; Chen and Pleskow 2007). A study investigating the use of cholangioscopy compared to cholangiography alone in PSC reported an improved sensitivity from 66% to 92% and specificity from 55% to 93% for the diagnosis of CC (Tischendorf, Kruger et al. 2006).

1.2.1.6 Endoscopic ultrasound and intra-ductal (miniprobe) ultrasound

Intra-ductal ultrasound probes were developed in order to improve imaging of bile wall thickening, presence of small stones and involvement of surrounding tissues for tumour staging. Studies report sensitivities of 89-91% and specificities of 50-80% for the diagnosis of malignant biliary strictures (Tamada, Ueno et al. 1998; Menzel, Poremba et al. 2000). EUS-guided fine needle aspiration cytology and biopsy may further improve accuracy in the diagnosis of malignant strictures (Eloubeidi, Chen et al. 2004).

1.2.1.7 Fluoro-2-deoxy-D-glucose, positron emission tomography (FDG-PET)

PET scanning uses radiolabelled compounds that incorporate into normal cellular pathways in tissues of interest. The fluorine-18 radiolabelled FDG

reagent is a hexokinase subtrate that is taken up and concentrated in cells with high levels of aerobic glycosylation such as tumour cells. The diagnostic accuracy of FDG-PET has been investigated in small studies. A retrospective study of 30 patients with indeterminate biliary strictures reported a sensitivity and specificity of 90% and 78% respectively for the diagnosis of malignant obstruction (Wakabayashi, Akamoto et al. 2005).

An early report suggested positron emission tomography (PET) scanning may be useful for surveillance or investigation of suspected cholangiocarcinoma in patients with PSC (Keiding 1998). However, a further study of 36 patients with suspected cholangiocarcinoma demonstrated a sensitivity of 85% for mass forming tumours, 65% for metastases but only 18% for infiltrating tumours (Anderson, Rice et al. 2004). Others have also demonstrated a low sensitivity of PET and high false positive rates in the presence of cholangitis (Fevery, Buchel et al. 2005). The likely reasons for this are that the FDG substrate is not specific for cancer cells and is taken up by all metabolically active cells such as glial cells in the brain and activated leukocytes (particularly neutrophils and macrophages) in areas of inflammation or infection, such as cholangitis. Also, the ductal infiltrating subtype of some tumours may account for some of the lower sensitivity due to the lack of mass formation to concentrate the tracer detected during scanning. Therefore, PET is not routinely utilised for diagnosis or surveillance of cholangiocarcinoma in PSC.

1.2.2 Histology and cytology

Pathological diagnosis using cytology or histopathology are required for confirmation of malignancy and so remain crucial in the investigation of biliary strictures. One study retrospectively assessed 185 patients undergoing major surgical resections for presumed malignant bile duct strictures and reported that 17% were benign on assessment of resected tissue, the alternative primary diagnosis being IgG4 disease (Erdogan, Kloek et al. 2008). When mass lesions are identified on cross sectional imaging, percutaneous biopsy for histopathology may be used and has a specificity > 95% for the diagnosis in BTC. The sensitivity has been little reported but is likely to be approximately 50% -70%. However, the procedure is uncomfortable for patients, may cause tumour seeding in potentially resectable disease and has risks including bleeding and biliary leak or perforation of neighbouring structures. The accuracy of percutaneous biopsy for the diagnosis of BTC is poorly reported in the literature but suggest tissue confirmation rates of 29-49% in larger modern series (Mansfield, Barakat et al. 2005; Witzigmann, Berr et al. 2006; Connor, Barron et al. 2007). The sensitivities for biopsy diagnosis are not reported in these series. In our own retrospective review of 105 cases at UCH (appendix), the sensitivity of percutaneous biopsy for pathological confirmation of BTC was 71%, with a specificity of 100% (Mills, Chapman et al. 2009).

The most commonly used method for gaining confirmation of malignancy is biliary brush cytology taken at the time of ERCP. However, the accuracy of this technique depends greatly on tissue sampling and processing techniques

as well as expertise of the cytopathologist (de Bellis, Sherman et al. 2002). Cytological interpretation and diagnosis is based on morphological changes which are partly subjective. One commonly used classification describes five categories; unsatisfactory, negative, atypical, suspicious and positive for malignancy (Logrono and Waxman 2001). Findings of atypia are common in the presence of inflammation and biliary stenting and usually represent benign disease. Presence of suspicious high grade dysplastic change has a high likelihood of malignancy. A review of 230 cases of bushings for biliary strictures in patients with PSC reported that suspicious change has a sensitivity of 46% and specificity of 97% for the diagnosis of malignancy (Charatcharoenwitthaya, Enders et al. 2008).

Biliary brush cytology sampling is a relatively straightforward procedure performed at the time of ERCP. A wire guided endobiliary brush is repeatedly passed across a stricture under the guidance of contrast radiography. The brush is removed and smeared across several microscope slides and then fixed using alcohol or preparatory cytology fixatives such as Cytofix™ and PreservCyt™. In some centres, brushes are cut off into saline or liquid fixatives and processed as ‘wet samples’ spun down in the laboratory and the pellet used for spreading on slides for assessment.

The specificity of brush cytology is very high and often reaches 100% (range 90-100%). However, there is a very wide range of reported sensitivity which is often disappointingly low (range 18-69%) (see Table 1 for a summary of results). One study suggests that therapeutic manipulation of strictures (such

as dilatation) before sampling may increase the sensitivity from 57% to 85% (Farrell, Jain et al. 2001).

The reported sensitivity of brush cytology in detecting CC in PSC is between 30 and 73% even when high grade dysplasia is included as a marker of underlying adenocarcinoma (Ponsioen, Vrouenraets et al. 1999; Khan, Davidson et al. 2002; Boberg, Jebsen et al. 2006). The sensitivity of cytology from bile aspirates is even lower. Other markers such as the presence of aneuploidy in biliary brushings or Kras and p53 in bile have been evaluated but are not sufficiently sensitive nor specific to be useful for screening or diagnostic tests (Kubicka, Kuhnel et al. 2001; Moreno Luna and Gores 2006).

Table 1 Biliary brush cytology for the diagnosis of malignant biliary strictures

Reference	n	Sensitivity (%)	Specificity (%)
(Foutch, Kerr et al. 1991)	30	33	100
(Kurzawinski, Deery et al. 1993)	93	69	100
(Lee, Leung et al. 1995)	149	35	100
(Pugliese, Conio et al. 1995)	94	54	97
(Ponchon, Gagnon et al. 1995)	204	35	100
(Mansfield, Griffin et al. 1997)	43	42	100
(Glasbrenner, Ardan et al. 1999)	78	56	90
(Macken, Drijkoningen et al. 2000)	106	57	100
(Jailwala, Fogel et al. 2000)	133	30	100
(Harewood, Baron et al. 2004)	113	24	100
(Baron, Harewood et al. 2004)	100	18	98
(Mills, Chapman et al. 2009)	96	31	100

Endobiliary biopsy may be performed by passing a biopsy forceps into the bile ducts and sampling tissues in areas of interest identified by contrast radiography. Reports of complications are higher than brush cytology but are

still low. Specificity reaches 100% and sensitivity is reported to range between 43-81% (Kubota, Takaoka et al. 1993; Ponchon, Gagnon et al. 1995; Pugliese, Conio et al. 1995; Sugiyama, Atomi et al. 1996; Schoefl, Haefner et al. 1997; Jailwala, Fogel et al. 2000). Our series of endobiliary biopsies (n=36) (Appendix) had a lower sensitivity of 41% but specificity of 100% (Mills, Chapman et al. 2009). EUS guided fine needle aspiration cytology and needle biopsy may also increase sensitivity but are more commonly applied to investigation of pancreatic pathology. Two small studies in CC (n=28 and 44) suggest that the sensitivity for diagnosis was increased to 86-89% by the use of EUS (Eloubeidi, Chen et al. 2004; Fritscher-Ravens, Broering et al. 2004). However, another study in CC suggested that EUS guided FNAC does not improve sensitivity (46%) in the absence of identifiable mass lesions (Rosch, Hofrichter et al. 2004).

Aspiration bile cytology is the simplest and safest method of obtaining samples for cytological assessment. Like biliary brushings, bile cytology has a very high specificity (reaching 100%) but has a lower sensitivity of 6-32% (see Table 2). Attempts to improve sensitivity by disruption of the stricture by dilatation prior to bile aspiration, may increase sensitivity from 29% to 63% (Mohandas, Swaroop et al. 1994). Other studies were unable to confirm this result and suggest that addition of bile aspiration cytology adds little to diagnostic accuracy if biliary brush cytology has been used (Kurzawinski, Deery et al. 1993; Mansfield, Griffin et al. 1997).

Table 2 Sensitivity of bile aspiration cytology for the diagnosis of malignant biliary strictures

Ref	n	Sensitivity	Specificity
(Desa, Akosa et al. 1991)	80	30	100
(Foutch, Kerr et al. 1991)	30	6	100
(Davidson, Varsamidakis et al. 1992)	62	30	100
(Kurzawinski, Deery et al. 1993)	93	32	100
(Sugiyama, Atomi et al. 1996)	43	32	100
(Mansfield, Griffin et al. 1997)	43	15	100

1.2.2.1 Chromosomal aberrations

Chromosomal aberrations are common in cancers. A study investigating the use of fluorescence in situ hybridisation (FISH) and digital image analysis (DIA) for detection of aneuploidy, tetraploidy and polysomy in suspicious biliary strictures reported a sensitivity and specificity of 88% and 63% respectively (Charatcharoenwitthaya, Enders et al. 2008). In PSC, the presence of any chromosomal change in biliary brushings had a sensitivity and specificity of 63% and 90%, respectively, for the diagnosis of BTC (Charatcharoenwitthaya, Enders et al. 2008).

1.2.3 Biomarkers and Tumour markers

A biomarker is any measurable biological parameter that alters with a disease state. These have historically been mostly protein products such as peptides identified by antibodies in ELISA tests or by immunohistochemistry but more recently the advent of high throughput technologies such as genomics, proteomics and metabolomics has opened up the possibilities for more complex biomarkers based on patterns of multiple gene expressions as biomarkers in the future. Biomarkers are often categorised as:

1. Diagnostic markers; markers that are associated with the presence of a disease. Eg, the presence of serum anti-mitochondrial antibodies are strongly associated with primary biliary cirrhosis.
2. Prognostic markers; markers that are associated with predicting outcome from a disease. Eg, predicting early or late death from an untreated cancer.
3. Predictive markers; markers that predict a response to an intervention. Eg, use of KRAS mutation analysis to determine the likely response to EGFR blocking drugs in colorectal cancer.

Tumour markers are measurable molecular markers that aim to identify the presence of cancers but few tumour markers to date are cancer specific. Most are protein antigens that have been identified using antibody studies. These can be diagnostic, prognostic and/or predictive. A number of tumour markers are considered to be very useful and remain widely used in routine clinical practice. These include prostate specific antigen (PSA) in prostate cancer, carcinoembryonic antigen (CEA) in colorectal cancer and carbohydrate

antigen 125 (CA-125) in ovarian cancer. Over recent years there has been a drive to use high throughput techniques such as proteomics and genomics to screen diseased tissues in order to identify new biomarkers. These techniques show great promise (Perou, Sorlie et al. 2000; Dhanasekaran, Barrette et al. 2001; Garber, Troyanskaya et al. 2001; Iacobuzio-Donahue, Maitra et al. 2002).

Sensitivity, specificity and positive predictive calculations using biomarkers.

Clinical applicability of biomarkers depends on a number of factors such as how common the condition is, whether they are required for screening, the importance of not making an incorrect diagnosis, ease of testing (both for patients and laboratories) and cost. For example, a screening test requires a very high sensitivity so that most or all tests are positive in the presence of a disease in order that clinical cases are not missed. Screening tests with a high sensitivity may suffer from lower specificity- ie the positive test may not be related to the disease of interest. For example, random blood glucose testing may be a good screening test for identifying people with diabetes mellitus but can have false positive results if tested soon after meals. The specificity can be improved by using fasting blood glucose which if abnormal is much more specific but can lose sensitivity in early disease. In cancer, a high sensitivity is important so that disease is not missed but also, a very high specificity is crucial to avoid mis-diagnosis of cancer with the potential for severe distress and inappropriate treatments being delivered to patients.

The positive and negative preditictive index (or precision rate) is the likelihood that a disease is present or absent based on measurement of a biomarker. A positive or negative predictive value (PPV/NPV) of 99% would make the likelihood of a disease being present or absent being very high or low respectively and so aid clinical decision making about the need for further confirmatory or excluding tests. Calculations for determining sensitivity, specificity and predictive indices are made using standard calculations below, measured against a gold standard for diagnosis (eg histopathology).

- Sensitivity= $\frac{\text{Number of true positives}}{\text{Number of true positives + number of false negatives}}$

- Specificity= $\frac{\text{Number of true negatives}}{\text{Number of true negatives + number of false positives}}$

- PPV= $\frac{\text{Number of true positives}}{\text{Number of true positives + number of false positives}}$

- NPV= $\frac{\text{Number of true negatives}}{\text{Number of true negatives + number of false negatives}}$

Few tumour markers to date are sufficiently specific for the confirmation of cancer so most cancers still require a pathological confirmation of diagnosis using cytology or histopathology. An exception is an elevated AFP in the presence of supportive imaging which, in combination, is considered sufficient for the diagnosis of hepatocellular carcinoma without biopsy confirmation. However, most tumour markers are currently used for supportive evidence of early diagnosis or for prognostication and predictive use, eg monitoring CEA in colorectal cancer to demonstrate a response to chemotherapy and/or

providing evidence of disease recurrence, or measuring PSA to identify patients becoming resistant to hormonal treatment in prostate cancer.

Currently there are no tumour markers specific for BTC. Here we summarise tumour markers that have been investigated or are used for the diagnosis and prognosis in BTC.

1.2.3.1 Liver biochemistry

Liver biochemical tests are invariably abnormal at presentation of BTC. However, any cause of biliary obstruction, cholestasis, cholangitis or hepatitis can cause abnormal liver biochemistry and so liver biochemical tests have a very low specificity for BTC. They are useful mainly for monitoring severity of biliary obstruction and response to treatments such as biliary drainage.

1.2.3.2 Carbohydrate Antigen 19-9 (CA 19-9)

Carbohydrate antigen 19-9 (CA19-9) is the most established tumour marker used in the diagnosis and follow up of patients with suspected BTC. CA19-9 is a complex glycoprotein first described in 1979 (Koprowski, Steplewski et al. 1979). The primary epitope is a sialylated carbohydrate component of the sialyl Lewis^{a} antigen, which is not produced by approximately 7% of the population and thus, CA19-9 is undetectable in these individuals (Steinberg 1990). Published series of patients with benign and malignant biliary obstruction report sensitivities and specificities of 50% to 90% and 54% to

98% respectively (Table 3) with a positive predictive value of 56% in one off the larger studies (Levy, Lymp et al. 2005). Many of the published series were in the setting of primary sclerosing cholangitis which accounts for a minority of cases of BTC and the utility of CA19-9 as a tumour marker in sporadic disease is less clear. It is also well recognised that any cause of benign biliary obstruction or cholangitis can cause falsely elevated levels of CA19-9 which falls following relief of obstruction or treatment of cholangitis (Bjornsson, Kilander et al. 1999; Madonia, Aragona et al. 2007).

Table 3 Accuracy of serum CA 19-9 for the diagnosis of BTC

Reference	CA19-9 threshold (U/ml)	n (CC)	Sensitivity	Specificity	Comments
(Chalasani, Baluyut et al. 2000)	100	26	75	80	CA19-9 only available in half cases. Addition CEA not helpful. PSC study
(John, Haghighi et al. 2006)	35	68	78	76	CA19-9 not associated with operability or prognosis. Higher CA 19-9 levels improved specificity but lower sensitivity
(Nichols, Gores et al. 1993)	100	9	89	86	Includes cohort described in later series from same unit (Levy 2005)
(Levy, Lymp et al. 2005)	63 129	14	90 78.6	98 98.5	Patients with elevated CA19-9 were all inoperable
(Bjornsson, Kilander et al. 1999)	37	9	63	85 (in combination with CEA)	High false positive rate in patients with cholestasis. No benefit of measurement in bile.
(Ramage, Donaghy et al. 1995)	200	15	60	90	Also used CA19-9 +(CEAx40) >400 with sensitivity of 66% and specificity of 100%
(Hultcrantz, Olsson et al. 1999)	37	4	Not reported	Not reported	Prospective study. Insufficient numbers to calculate sensitivity and specificity but suggest both low
(Patel, Harnois et al. 2000)	100	103	53	76-92	Sporadic BTC
(Fisher, Theise et al. 1995)	74	3	50	54.5	PSC patients undergoing OLT
(Burak, Angulo et al. 2004)	180	44	66.7	97.7	PSC study
(Siqueira, Schoen et al. 2002)	180	55	67	98	PSC study
(Lindberg, Arnelo et al. 2002)	100	57	67	89	PSC study
(Kim, Kim et al. 1999)	37	322	73	63	PSC study
(Buffet, Fourre et al. 1996)	40	308	92	72	Biliary and pancreatic cancer

A further characteristic of CA19-9 is that it tends to be elevated predominantly in advanced disease. Ideally, a useful tumour marker would be able to detect early, treatable disease but reports suggest that significantly elevated levels (eg.>129 U/ml determined by ROC curves) are only present in inoperable cases (Charatcharoenwitthaya, Enders et al. 2008). Cha et al studied 61 surgically resected 'early bile duct cancer' patients ("carcinoma with invasion confined within the fibromuscular layer of the extra-hepatic bile duct") and reported a CA19-9 sensitivity as low as 33% with the threshold set at >37U/mL and of only 15% when the CA 19-9 threshold was raised to >100U/mL (Cha, Kim et al. 2006). In addition, CA 19-9 levels are also raised in many other malignant conditions, including, pancreatic cancer (sensitivity, 68% - 94%, specificity 76% - 100%), colorectal, gynaecological and gastric carcinomas (Rhodes 1999).

Measurement of CA19-9 has been advocated for the surveillance of CC in patients with PSC (Table 4). However, despite many clinicians using this tumour marker in clinical practice, there is little evidence that it facilitates early diagnosis or alters outcome in such patients (Levy, Lymp et al. 2005).

Table 4 Accuracy of CA19-9 for the diagnosis of BTC in PSC, using a selection of low and high cut-off levels. (Reproduced from Charatchoenwitthaya et al 2008)

	CA 19-9 cut-off level (U/ml)				
	20	**40**	**100**	**129**	**200**
Sensitivity (%)	78	57	22	13	13
Specificity (%)	67	84	99	100	100
PPV (%)	23	30	71	100	100
NPV (%)	96	94	91	90	90

Measurement of CA19-9 in bile in addition to measurement of serum levels does not improve the accuracy for the diagnosis of BTC (Ker, Chen et al. 1991).

1.2.3.3 Carcinoembryonic antigen (CEA)

CEA, like CA19-9, is a complex glycoprotein and is commonly used for monitoring response to treatment and disease recurrence in colorectal cancer (Duffy 2001). CEA has a reported sensitivity of 33 to 68% and specificity of 81 to 86% for the diagnosis of BTC (Ramage, Donaghy et al. 1995; Bjornsson, Kilander et al. 1999; Siqueira, Schoen et al. 2002). The combination of CA19-9 and CEA has been advocated as a more reliable tumour marker but still has a reported sensitivity and specificity of only 33% and 85%, respectively (Bjornsson, Kilander et al. 1999).

1.2.3.4 Carbohydrate antigen-125 (CA-125)

The reported sensitivity and specificity for carbohydrate antigen 125 (CA 125) for the diagnosis of cholangiocarcinoma are <70% and 76% (Chen, Shiesh et al. 2002). As with the other tumour markers, it may be falsely elevated in other cancers (particularly ovarian), peritoneal disease and cirrhotic ascites (Miralles, Orea et al. 2003).

1.2.3.5 Mucin 4 (MUC4) and mucin 5AC (MUC5AC)

Epithelial surfaces express tissue specific patterns of mucins which may change in neoplasia. A study from our group reported that biliary MUC4 and serum MUC5AC had a sensitivity of 58% and specificity of 87% for the diagnosis of BTC. The presence of MUC5AC in serum was also a marker of poor prognosis (Matull, Andreola et al. 2008). However, patients with PSC commonly had elevated levels of these mucin biomarkers, so reducing their value as biomarkers for malignancy in this patient population.

1.2.3.6 Minichromosome maintenance markers (mcm)

DNA replication during the process of cell division is tightly regulated in the cell cycle by a number of inhibitory and stimulatory proteins. Mcm 2-7 are a group of proteins that combine with other 'licensing' proteins, such as the ORC (origin recognition complex) and CDC6, onto chromatin as part of the pre-replicative complex that initiates DNA replication. They are expressed during the cell cycle in actively proliferating cells such as cancer cells but are down-regulated in non-replicating, mature, differentiated cells. Important proteins inhibiting the process of DNA replication include Cdk1, Cdk2 and geminin which are bound and inactivated by other proteins such as APC/C. Over-expression of mcm's is an early feature of cancer and these mcm's have been shown to be shed from epithelial surfaces and act as markers for other malignancies such as bladder cancer (Stoeber, Swinn et al. 2002). A study by our group has demonstrated that measurement of mcm5 in bile samples (n=102) is a marker of malignancy and has an improved sensitivity (66%

versus 20%) and similar specificity (94%-100%) to biliary brush cytology of indeterminate biliary strictures (Ayaru, Stoeber et al. 2008). This is supported by a much higher rate of immunohistochemical staining for mcm2 and mcm5 in tissue sections from patients with malignant versus benign biliary strictures (median 76.5% mcm2 positivity for malignant strictures and 5% in benign disease, $P<0.005$) (Ayaru, Stoeber et al. 2008). A larger, multicentre, prospective study on the use of mcm's as diagnostic markers in pancreobiliary cancer is planned.

1.2.3.7 Other circulating tumour markers

There are limited data on a number of other potential biomarkers in the diagnosis and prognosis of BTC. These include a small prospective trial assessing the utility of CA-242 and CA50 in patients with PSC and suggest poor sensitivity and specificity and are not recommended for use (Hultcrantz, Olsson et al. 1999). CK19 represents a promising method of detecting circulating tumour cells in patients with BTC or other adenocarcinoma's (Takada, Masuda et al. 1995; Brechot, Chevret et al. 1997; Nakata, Takashima et al. 2004; Pujol, Molinier et al. 2004; Andreadis, Touloupidis et al. 2005) and we have further investigated the role of CK19 fragments (CYFRA 21-1) in chapter 2.

1.2.3.8 Genetic markers of BTC

Mutations in the K-RAS gene are described in many cancers including BTC. Studies assessing the accuracy of K-RAS mutations in biliary brushings report

sensitivities of 27-64% and specificities of 62-100% (Saurin, Joly-Pharaboz et al. 2000; Wang, Yamaguchi et al. 2002). Mutations in the tumour suppressor gene p53 have also been investigated with a prevalence of 52% in cholangiocarcinoma (Liu, Zhang et al. 2006).

Other potential genetic markers in BTC have been investigated, most commonly using immunohistochemistry in surgical specimens. These include p16, c-MET and telomerase (Morales, Burdick et al. 1998; Klump, Hsieh et al. 2003; Enjoji, Nakamuta et al. 2004). None have good sensitivity for diagnosis.

1.2.4 Conclusions: Diagnosis of BTC

BTC remains challenging to diagnose, particularly in conditions with pre-existing or predisposing complex biliary stricturing such as PSC. Currently, cytopathology and/or histopathology remain the only definitive methods for diagnosis. However, these involve invasive procedures with significant risks and the sensitivities for such tests are low (range 18%-69% for biliary brush cytology).

Although some non-invasive tests such as MRCP or serum CA19-9 levels are relatively useful in diagnostic algorithms, none have sufficient accuracy to be reliable tests for diagnostic or screening purposes. Recent advances in surgical and palliative treatment mean that more patients may be suitable for therapy in the future. However, these approaches require a definitive diagnosis prior to therapy, particularly with reports of high rates of benign disease in patients undergoing major hepatic resections for presumed CC (Erdogan, Kloek et al. 2008). New improved biomarkers are needed to improve the clinical care of patients with BTC, and this was the overall rationale for the thesis.

1.3 Approaches to biomarker discovery.

There are a number of potential approaches to biomarker discovery. Most methods are based on the principle of detecting differences between normal and diseased tissues. Many of the tumour markers in current use, including AFP, CA19-9 and CA125, were discovered by screening of poyclonal antibodies raised against different tissues and disease states. Some of these antibodies bound with relative specificity to cancer tissues. The antigenic components were later identified and improved antibodies to those antigens have been produced and used in protein detection assays such as tumour marker ELISA tests. The current most widely used methods for biomarker discovery use high throughput techniques described below with the addition of complex computer software analysis programmes to screen the large quantities of data generated. Nomenclature for the methods used vary between authors and centres but formal definitions are being developed. Many of the methods and disciplines are complementary and may be carried out in parallel to gain different but inter-related information of biological systems, often with the support of bioinformatics software and expertise.

Genomics loosely refers to the application of genome scale technologies for biological investigation. Genomic studies include investigation of DNA mutations, DNA methylation status, single nucleotide polymorphisms, gene expression analysis by measuring mRNA expression, assessment of splice varients using exon arrays and investigation of non coding RNA molecules including microRNA. The term transcriptomics has been introduced to separate out investigation of the transcriptome, defined as the whole set of

RNA products produced by the genome at a particular point in time. The genome is essentially stable in time with the exception of development of new mutations and alteration in splice variants. However, the transcriptome is dynamic and changes with normal physiology and disease, depending on the relative abundance and activity of different RNA species including mRNA, rRNA, microRNA and other small interfering RNA species. We chose to use a transcriptomic approach to biomarker discovery for this project concentrating on gene expression measured by mRNA expression microarrays. The most commonly used high throughput assays are cDNA microarrays with complementary probes which span sequences of expressed mRNA of genes throughout the entire genome. These are indicators of gene expression rather than structural alterations in the source genes and are discussed in more detail in Chapter 5 using the example of mRNA expression arrays used in our work.

Other important genetic alterations are also crucial to cancer biology but are not further investigated in this thesis. These include DNA mutations where, depending on the nature of the mutation, a mutated gene may or may not be transcribed or transcribed to an abnormally functioning protein. For example, point mutations in the KRAS gene can be implicated in various gastrointestinal cancers where an abnormal variant of the protein is translated but this is functionally altered so that it is activated in an uncontrolled fashion independently of the many usual upstream activators such as activation of the EGFR receptor complex.

Proteomics is another field of research commonly used for biomarker discovery. It involves the large scale study of protein expression and function. A potential advantage over genomic studies is that the protein products being investigated are at later stages in biological pathways and diseases of interest. Although siginifiant post translational modification of proteins occurs, this is less significant than the large scale post transcriptional modification that occurs in gene expression pathways. Protein modifications, such as protein phosphorylation and methylation, are crucial to the normal functioning of proteins and may differ in disease. Biomarker discovery using proteomics aims to identify alterations in disease states. The most widely used techniques include 2D gel electrophoresis, high pressure liquid chromatography and mass spectroscopy assays including MALDI and SELDI-TOF. Our group, in collaboration with Dr John Timms (Cancer Proteomics Laboratory, Institute for Women's Helath, UCL), is investigating the use of proteomics for biomarker discovery in BTC alongside the genomic studies described in this thesis. Data from these proteomic studies are not described in this thesis. Similar studies in pancreobiliary cancer have been published in recent years and provide an additional source of data on proteins that may be further investigated. Examples include Annexin IV, Golgi Membrane Protein 1, AGR2 and S100A10, (Kristiansen, Harsha et al. 2008; Sitek, Sipos et al. 2009).

Other approaches used less commonly include lipidomics (the study of lipids) and metabolomics (the study of metabolites). Like proteomics, both have

strength in biomarker discovery in that they measure end products of biological pathways which may identify alterations specific to disease. Both disciplines use complex large data generating techniques such as chromatography, mass spectroscopy and nuclear magnetic resonance. These methods are not described further in this thesis.

An alternative approach to high throughput screening techniques such as genomics and proteomics is to directly identify tumour cells in biological samples such as blood (circulating tumour cells). This approach is discussed in more detail in the introduction to Chapter 2 before concentrating on the main subject of whole genome expression analysis in the following chapters.

All of these methods are useful individually and increasingly together now that software allowing integration of data from multi-disciplinary techniques is being developed. These allow deeper investigation of the various “hallmarks of cancer” as outlined by Hanahan and Weinberg (Hanahan and Weinberg 2000). They described six key alterations in cell biology that are common to most or all cancer cells which are 1) self sufficiency in growth signals, 2) evading apoptosis, 3) insensitivity of anti-growth signals, 4) sustained angiogenesis, 5) tissue invasiveness and metastasis and 6) limitless replicative potential (Hanahan and Weinberg 2000).

Examples of these mechanisms of cancer biology in BTC are summarised in various reviews (Fava, Marzioni et al. 2007; Blechacz and Gores 2008; Sirica, Nathanson et al. 2008; Wise, Pilanthananond et al. 2008) and include:

- *Self sufficiency in growth signals and evading apoptosis.*

IL6 has been widely associated with the pathogenesis of BTC. This in part may be related to chronic biliary inflammation causing elevated levels of IL6 secreted from local inflammatory cells such as that seen in chronic *Clonorchis* infestation, one of the major risk factors for intrahepatic CC. However, cholangiocytes themselves also secrete IL6 acting in a paracrine loop. Elevated levels of IL6 inhibit apoptosis via stimulation of MCL1 and STAT3 promoting malignant proliferation (Isomoto, Kobayashi et al. 2005).

- *self sufficiency in growth signals*

MAP2K2 and MAP2K1 belong to the MAP kinase kinase family and phosphorylate and thus activate MAPK1/ERK2 and MAPK2/ERK3 stimulating cell proliferation. The MAPK pathway is commonly upregulated in BTC and KRAS mutations associated with increased MAPK activity is a common early finding in CC. Inhibition of the MAPK pathways using siRNA inhibits the growth of cholangiocarcinoma cell lines in vitro (Leelawat, Leelawat et al. 2006).

- *evading apoptosis*

Inactivating mutations in the tumour suppressor genes $p16^{INK4a}$ supporting evasion of cell death are reported in BTC such as in patients with PSC and PSC related CC. Other mechanisms include over-expression of the anti-apoptotic MCL1 in CC (Wise, Pilanthananond et al. 2008).

- *tissue invasiveness and metastasis*

Elevated expression of various matrix metalloproteinases (MMP2, MMP7, MMP9 and MT1-MP) that allow enhanced tissue invasiveness and abnormal cell anchoring via mutated cell adhesion molecules such as E-cadherin are

reported by DNA mutation analysis, mRNA expression profiling and tissue levels by immunohistochemistry in BTC (Fava, Marzioni et al. 2007). Other proteolytic enzymes that allow increased invasiveness such as PRSS1, PRSS2 and aspartyl b-hydroxylase are also up-regulated in CC and inhibition of the latter impairs the invasiveness of CC cell lines *in vivo* (Maeda, Sepe et al. 2003; Nakanuma, Tajima et al. 2010).

- *sustained angiogenesis*

Increased expression of enhancers of angiogenesis such as VEGFA and VEGFC are common in BTC (Fava, Marzioni et al. 2007).

- limitless replicative potential

BTC cells have elevated telomerase acitivity, which is also enhanced by elevated levels of IL-6 commonly implicated in the biology of BTC as above (Morales, Burdick et al. 1998; Yamagiwa, Meng et al. 2006).

1.4 Preview of thesis

The following chapters describe work done in order to identify and investigate potential novel biomarkers in BTC with a particular emphasis on gene expression analysis. Chapter 2 introduces the concept of circulating tumour cells and describes the measurement of circulating fragments of an epithelial cell marker cytokeratin 19 (CYFRA 21-1), as an example of the approach to the clinical application and testing of biomarkers with patient samples.

Chapter 3 is divided into two parts and investigates the suitability of formalin fixed paraffin embedded (FFPE) tissue for gene expression studies. Part 1 examines the use of whole tissue sections and Part 2 examines the use of laser capture microdissection of FFPE tissues. Results showed poor levels of RNA isolated from such tissues, and with the rarity of the PSC related CC samples, we elected to investigate the use, and demonstrate suitability of, endoscopic biliary brushings for gene expression profiling (Chapter 4).

Chapter 5 provides methodology and a results summary for whole genome RNA expression analysis of biliary brushings using microarray technology. These results are then validated in Chapter 6 with alternative methods for measurement of gene expression using quantitative real time polymerase chain reaction (qPCR). The final results chapter (Chapter 7) investigates whether a selection of genes identified in the earlier work are translated to protein products which may be later tested as potential biomarkers in clinical samples. The final section (Chapter 8) is a discussion of issues raised during the project and potential future work.

1.5 Aims of this project

The central aim of this project was to identify novel biomarkers with which to diagnose and predict prognosis of BTC, with a particular emphasis on cholangiocarcinoma.

Specific aims were:

- To investigate whether measurement of circulating fragments of cytokeratin 19 (CYFRA 21-1) may be a biomarker for BTC and to assess the suitablitiy of our biobank for later validation studies.
- To investigate whether archived explanted liver tissue from patients with PSC related CC and benign PSC disease may be used to identify genes as biomarkers or risk factors for development of BTC in patients with PSC.
- To investigate whether endoscopic biliary brushings may be used to detect altered gene expression in BTC
- To investigate whether aberrantly expressed genes identified in RNA expression studies are translated to altered protein expression in biliary tissues of patients with BTC.

Chapter 2.

2 Identification of circulating tumour cells in blood and use of CYFRA 21-1 as a biomarker of biliary tract cancer.

2.1 Background

One approach to diagnosing BTC is to identify malignant cells in circulating blood. Recent advances have allowed small numbers of circulating tumour cells (CTCs) to be detected in blood or other body fluids such as urine, bile, stool and pleural fluid in patients with malignant disease. However, these methods are time consuming and technically difficult. For these reasons, they were not yet widely available for research use when we began theses studies.

Another approach is to use tumour specific markers. However there are few such markers and most are also produced to some degree in normal tissues reducing their specificity. The tumour markers PSA, CEA and CA19-9 are up-regulated in some cancers but are also identifiable at low levels in normal tissues. Methods such as ELISA and PCR are often so sensitive that they may result in false positive tests in absence of malignancy with a normal threshold being difficult to define.

Under normal circumstances, there are few, if any, identifiable circulating epithelial cells in healthy individuals. Therefore, most assays for detecting circulating tumour cells in blood are based on the identification of normal, 'surrogate' epithelial cell markers (eg EpCAM, CK19) rather than tumour specific markers. Studies using such markers consistently demonstrate the absence of circulating epithelial cells in healthy controls (Cristofanilli, Budd et al. 2004; Fizazi, Morat et al. 2007; Nagrath, Sequist et al. 2007; Gervasoni, Monasterio Munoz et al. 2008). These markers have the advantage that they have a potential for population screening for multiple adenocarcinomas. The disadvantages of this method are the potential for false positive results, the lack of sensitivity in tumours of non epithelial cell origin, and the possibility that positive results will only be found in metastatic or advanced, inoperable cancers.

2.2 Methods used for detecting circulating tumour cells

2.2.1 PCR based identification of epithelial specific nucleic acids

These assays require the assumption that there is no epithelial cell specific nucleic acid in circulating blood. The majority use RT-PCR to identify mRNA for epithelial cell markers including various cytokeratins, mucins and CEA. To

improve the sensitivity and specificity of such tests, protocols often employ cell separation techniques to purify the samples prior to RT-PCR.

2.2.2 Cell separation using immuno-magnetic beads

Cell separation techniques are now commercially available from a number of manufacturers (e.g. Dynabeads®). They are based either on positive selection of the cells of interest using antibody coated magnetic beads, or by 'cell depletion', using beads to capture and remove contaminating cells (most commonly leukocytes using the leukocyte common antigen CD45), leaving the cells of interest in solution. Although these techniques are valuable to provide a higher yield, improvements in RNA extraction techniques mean that sufficient RNA can be extracted from as few as 5 cells.

This approach was used by Gervasoni et al (Gervasoni, Monasterio Munoz et al. 2008). Ficoll based fractionation of blood was followed by removal of leukocytes using CD45 immuno-magnetic beads, and then RT-PCR to identify epithelial cell markers including CEA, CK18, CK19, CK20, GCC, EpCAM, MUC1 and TERT. A similar method was used to measure circulating telomerase activity using a PCR-ELISA assay in patients with cancer of the prostate (Fizazi, Morat et al. 2007). The assay detected 79% of patients with advanced prostate cancer (n=24), 79% with early localised disease (n=70) and none of the healthy controls (n=22).

2.2.3 Cell separation using fluorescence activated cell sorting (FACS)

As techniques for identifying CTCs improve, these methods could potentially be automated using FACS machines. The main limiting factor at present is the efficiency of cell separation for complex clinical samples, such as blood.

2.2.4 'Blood on chip' devices

The term 'blood on chip' device has been used to describe semi-automatic systems developed to extract CTC's from blood. One such sensitive system is the EpCAM "CTC-chip" (Nagrath, Sequist et al. 2007). This uses a chamber filled with micro-posts coated with anti-Ep-CAM antibodies and uses image analysis software to identify trapped epithelial cells which can then be isolated and analysed. The sensitivity and specificity of this system was reported to be over 99% in 116 patients with adenocarcinomas of the colon, breast, prostate and lung. The system was also sensitive in a subgroup of 7 patients with early cancers. No CTCs were identified in any of the healthy controls (n=20). Systems such as this show promise for the development of screening tests for adenocarcinomas, the characterising of tumours to direct appropriate medical therapy, and the isolation of CTCs for more detailed biological study.

2.2.5 Commercial CTC assays.

Although these were not available at the initiation of this project, commercial assays such as the CellSearch™ assay (Quest Diagnostics) have recently become available. These use a combination of the methods described above

to isolate a purified sample of CTCs from peripheral blood. The CellSearch™ assay relies on sequential use of antibodies to EpCAM, CD45, and then a combination of CK8, CK18 and CK19 antibodies to isolate the CTCs. This assay has been applied to the investigation of other cancers. For example, in breast cancer, it has shown that the identification of CTCs using this technique predicted future metastatic disease and poorer overall survival in patients clinically staged as having non-metastatic disease (Bidard, Mathiot et al. 2009).

2.3 Measurement of circulating CK19 fragments (CYFRA 21-1) for the diagnosis and prediction of prognosis in patients with biliary tract cancer

2.3.1 Introduction

Of the CTC markers investigated in other cancers, the most suitable candidate biomarker for use in BTC may be cytokeratin 19 (CK19), which is a constituent of the intermediate filament proteins responsible for the structural integrity of epithelial cells. CK19 is constitutively expressed by many epithelial cells, is a highly sensitive cholangiocyte marker and is also commonly over-expressed by biliary tract cancer cells (Maeda, Kajiyama et al. 1996). As with other epithelial markers, CK19 is rarely detected in the blood of healthy individuals. Measurement of circulating CK19 fragments have been shown to be a biomarker for other malignancies including non small cell lung cancer (Takada, Masuda et al. 1995; Brechot, Chevret et al. 1997), bladder cancer (Andreadis, Touloupidis et al. 2005), breast cancer (Nakata, Takashima et al. 2004), and gastric cancer (Nakata, Chung et al. 1996). Circulating CK19 fragments have also been shown to be a marker of prognosis and of disease recurrence after surgery in non small cell lung cancer (Brechot, Chevret et al. 1997; Pujol, Molinier et al. 2004). These studies used the CYFRA 21-1 immuno-assay which utilises antibodies directed against CK19 fragments,

allowing detection of circulating fragments in blood, even after degradation of the intact protein.

CYFRA 21-1 was first reported to be a marker of BTC in a report of 4 cases of intrahepatic cholangiocarcinoma (Kashihara, Ohki et al. 1998). A more recent study reported that CYFRA 21-1 had a sensitivity and specificity of 75% and 92% respectively for the diagnosis of CC and that a low CYFRA 21-1 (<2.7 ng/ml) was a strong predictor of disease free survival after attempted curative surgical resection with 5 year survival rates of 76% v 25% in those with levels below or above 2.7 ng/ml respectively (Uenishi, Yamazaki et al. 2008). These studies were done in Japanses patients with intrahepatic cholangiocarcinoma, a disease which is distinct from the extrahepatic cholangiocarcinoma and gall bladder cancer seen more commonly in the Western world. No data has been published in such patient populations.

The current study was designed to evaluate the accuracy of CYFRA 21-1 in the diagnosis and prediction of outcome of biliary tract cancer and to assess whether this method may have a role in screening for biliary tract cancer in high risk patients such as those with PSC.

2.3.2 Materials and Methods

2.3.2.1 Patient population and clinical samples

The study was conducted following ethical approval granted by the Joint UCL/UCLH Ethics Committee (reference 06/Q0152/106). Blood samples (n=124) were collected prospectively using Vacutainer tubes (BD, New Jersey, USA) from patients with;

- benign biliary disease (n=58);
    - PSC [n=19]
    - papillary stenosis or sphincter of Oddi dysfunction [12]
    - choledocholithiasis [7]
    - stricture secondary to chronic pancreatitis [8]
    - inflammatory stricture [2]
    - autoimmune pancreatitis [7]
    - healthy controls [3],
- Biliary tract cancer (n=66)
    - BTC [n=60; cholangiocarcinoma (56), gallbladder cancer (4)]
    - PSC-related cholangiocarcinoma [n=6]

Blood samples were separated by centrifugation at 2500rpm for 8 minutes and stored at -80^{0}C until further analysis.

Baseline patient characteristics and routine blood tests including liver biochemistry, blood count and inflammatory markers, were recorded at the time that blood was taken for CYFRA 21-1 measurement. In 120 of the 124

patients, CA19-9 was also measured. All patients in the cancer group had cytological or histologic confirmation of malignancy consistent with biliary tract cancer. PSC was diagnosed on the basis of well described clinical characteristics, primarily based on cholangiographic findings, and where possible, supportive evidence such as the presence of IBD or liver histology consistent with the diagnosis (Chapman 1980). None of the patients in the benign group had evidence of malignancy after a minimum of 6 months follow up (median 12 months). Cancers were staged using the TNM and the American Joint Committee on Cancer staging systems (American Joint Committee on Cancer Staging, New York, Springer Verbay 2005). T1/T2 disease without evidence of nodal or metastatic spread was classified as 'early' disease (stage I/II). The remainder (stages III/IV) were classified as 'advanced' disease. Median follow up from the date of blood collection to the date of death or the study end date was 12.9 months (range 7.2-70) in the benign/PSC group and 6.7 months (range 0.3-40) in the BTC group. Survival time was calculated from the date of blood sampling.

All patients in the benign group were alive at the end of the study period and only one had surgery for a complex biliary stricture related to autoimmune pancreatitis. Three patients died in the PSC group (two with advanced PSC, one with cardiac disease) and a further three underwent liver transplantation during the follow up.

Table 5. Baseline patient characteristics for blood samples used in the CYFRA 21-1 analysis

	Benign disease n=39	PSC n=19	PSC/CC n=6	BTC n=60
Age (years)	53 (25-80)	50 (20-75)	60 (25-73)	68 (34-91)
Gender (M:F)	21:19	10:9	4:2	30:30
Bilirubin (μmol/L)	14 (5-335)	17 (7-341)	81 (28-375)	43 (8-676)
CA19.9 (U/ml)	10.5 (0-686)	13.5 (0-3145)	473 (129-4,139)	316 (0-145,528)

2.3.2.2 Measurement of CYFRA 21-1 and CA19-9

Measurement of serum or plasma CK19 fragments was performed using the CYFRA 21-1 ELISA kit (DRG International, Marburg, Germany) according to the manufacturer's instructions. The assay uses 2 mouse monoclonal antibodies (KS19.1 and BM19.21) for the detection of CK19 fragments. 50μl of sample were used in duplicate on 96 well ELISA plates. Horseradish peroxidase was used for the colour immunometric assay. Absorbance was measured at 450nm with the mean concentrations per pair calculated using linear correlation. Serum samples were used for measurement of CYFRA 21-1 where possible [n=86]. In others [n=38], only EDTA plasma samples were available. According to the manufacturer, EDTA plasma may give falsely elevated CYFRA 21-1 results. In order to test the correlation between serum and plasma results, we measured CYFRA-21-1 levels in paired serum and plasma samples from 13 patients. Using both the Pearson correlation and paired T test tests, plasma levels had a mean level 1.3 times the level of serum CYFRA 21-1 (R^2= 83.3, p=0.001) (Figure 3). Serum and corrected plasma levels were used for data analysis. Bilirubin does not impact on

measurement of CYFRA 21.1 levels (personal communication, DRG Inernational, Marburg, Germany).

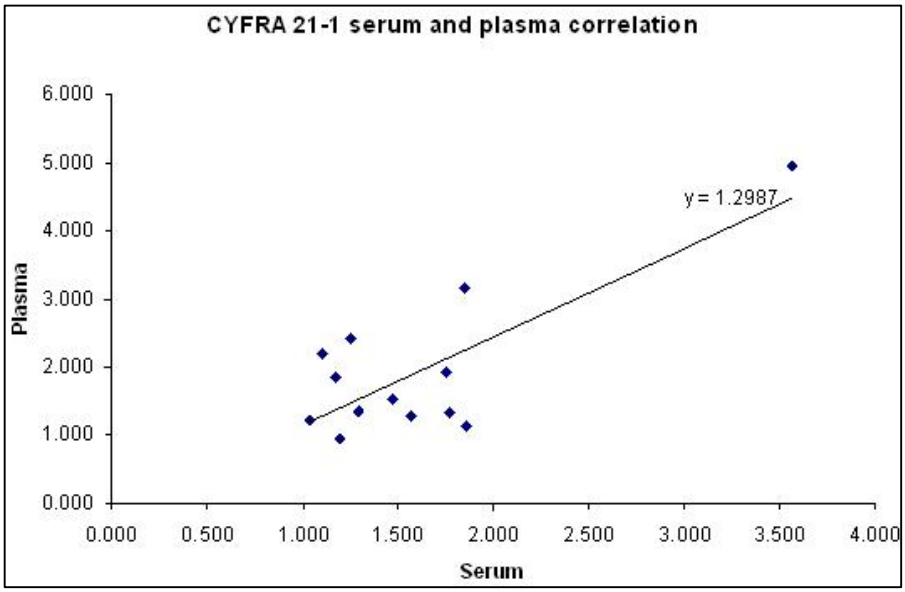


Figure 3. Correlation between serum and plasma CYFRA 21-1 using paired samples from the same patients showing a higher plasma CYFRA 21-1 requiring a correction factor of 1.3 for comparison with serum levels

Analysis of CYFRA 21-1 levels in combination with CA19-9 was also made using an optimal cut-off for CYFRA 21-1 of 1.5ng/ml as determined by the ROC curve, and a CA19-9 level of 37 U/ml as the limit of normal range used by our laboratory. Analysis of the combination of both elevated biomarkers was made for benign and malignant disease groups. In line with published data for CA19-9 and CYFRA 21-1, we also assessed the value of both markers with higher cut-off levels as a predictor of prognosis.

2.3.2.3 Statistical analysis

Statistical analysis was performed using SPSS software version 14.0. Comparison of groups was assessed using the Kruskall Wallis and Chi square tests. Survival data are presented using the Kaplan-Meier method with comparison made using the log rank test. Statistical significance was set at a p value of 0.05 for all tests. The area under the curve method was used to assess the efficacy of the biomarkers and to determine a suitable cut-off level for CYFRA 21.1 in order to calculate sensitivity and specificity.

2.3.3 Results

Baseline patient characteristics are listed in Table 5. Patients with benign disease had lower median CA19-9 and bilirubin levels, which were highest in the PSC related CC group. CYFRA 21-1 and CA19-9 levels are shown in Figure 4.

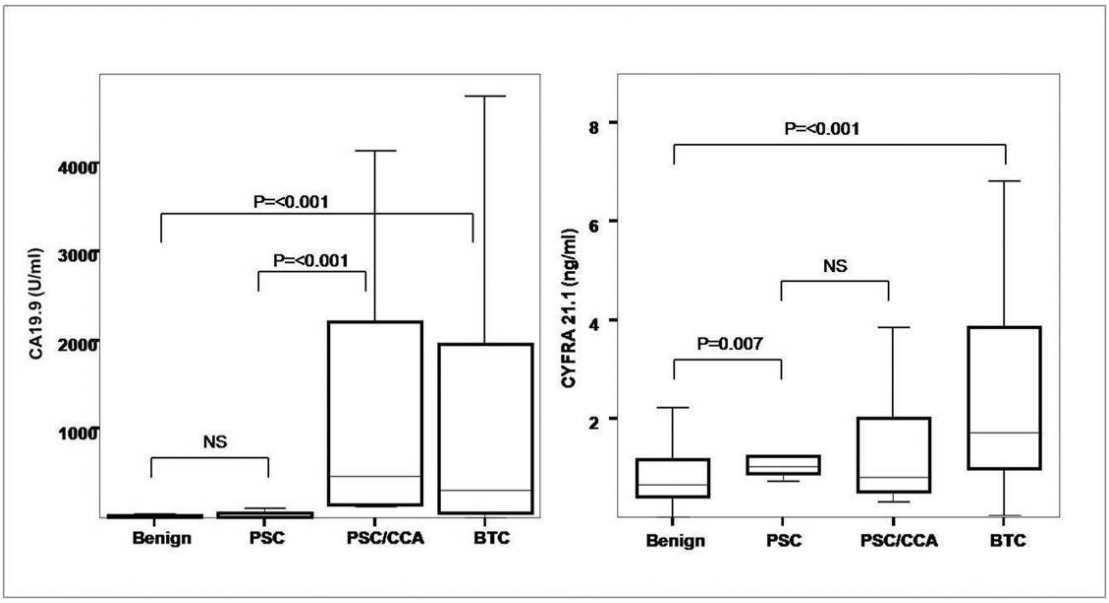


Figure 4. Median levels for CA19-9 and CYFRA 21-1 for the different patient groups with 75th centiles shown by the bars. Both CA19-9 and CYFRA 21-1 had significantly higher levels in BTC compared to benign controls (PSC and non PSC controls).

2.3.3.1 Efficacy of CA 19-9 and CYFRA 21-1 as tumour markers for BTC.

To analyse the efficiency of CA 19-9 and CYFRA 21-1 as tumour markers for BTC, comparisons were made between all patients with cancer, against all patients with benign disease, including those patients with PSC. Using a cut-off for CA 19-9 of ≥37 U/ml, sensitivity, specificity, PPV and NPV were 79%, 78%, 81%, and 75% respectively. These figures are in keeping with published figures. The optimal cut-off for CYFRA 21-1, determined using the ROC curve

method, was ≥1.5ng/ml (Figure 5). At this level, CYFRA 21-1 had a sensitivity, specificity, PPV and NPV of 56%, 88%, 84% and 64% for the diagnosis of BTC. Data were also calculated using higher cut-off figures for CA19-9 (129U/ml) and CYFRA 21-1 (3.0ng/ml) (Table 6).

When using serum samples alone (n=86), results for determining sensitivity and specificity were similar (as was an optimal cut-off of 1.5ng/ml), but had a lower statistical significance (sensitivity 53%, specificity 86%, AUC 68%, p=0.004 v 0.001). Using a cutoff of ≥3 ng/ml with only serum samples, the sensitivity was 25% and specificity 95%.

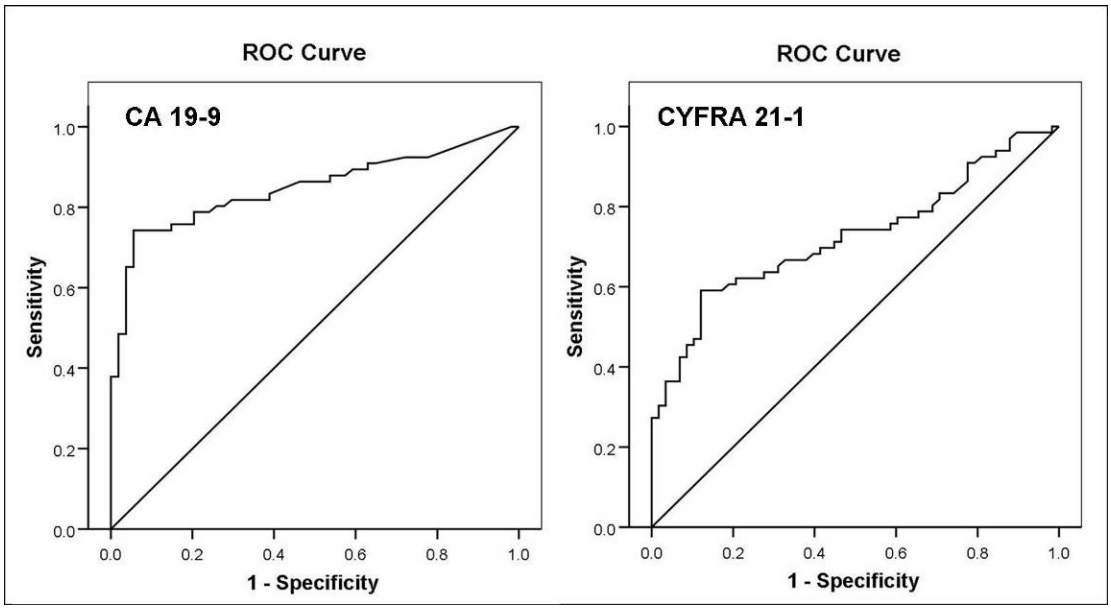


Figure 5. Receiver operating characteristics (ROC) curves for CA19-9 (AUC 84% [77-91%]) and CYFRA 21-1 (AUC 73% [64-82%]).

Table 6. Efficacy of CA19-9 and CYFRA 21-1 tumour markers in the diagnosis of BTC

	CA 19-9 >37	CA 19-9 >129	CYFRA 21-1 >1.5	CYFRA 21-1 >3.0	CA19-9 >37 + CYFRA 21-1 >1.5	CA19-9 >129 + CYFRA 21-1 >3.0
Sensitivity	79	74	56	30	45	27
Specificity	78	95	88	97	96	100
PPV	81	94	84	91	94	100
NPV	75	76	64	55	59	53

2.3.3.2 Combination of CYFRA 21-1 and CA 19-9 in the diagnosis of BTC.

In order to assess the role of these markers in combination, results were grouped into those positive for CA 19-9 and CYFRA 21-1 (+/+) [n=32], those positive for one or other marker (+/-) [n=43], and those negative for both (-/-) [n=45]. To calculate sensitivity and specificity of positive (+/+) tests, the results of +/- and +/- were pooled into one group. Thus the comparison was between +/+ and combined +/- & -/-. Data from the lower and higher cut-off figures (37U/ml & 129U/ml for CA19-9, and 1.5 ng/ml & 3.0 ng/ml for CYFRA 21-1) were grouped in the same way. The calculated sensitivity, specificity, PPV and NPV are shown in Table 6. Of note, in only 2 patients with benign disease were both markers positive using the lower cut-off values. None were positive for both using the higher cut-off values.

2.3.3.3 CYFRA 21-1 as a marker of tumour stage and prognosis in BTC.

Patients with advanced disease (stages III-IV) had significantly higher CYFRA 21-1 levels (median 2.42 ng/ml, range 0.49 -35) than those with early stage I or II disease (1.03 ng/ml, range 0.32 to 5.59) (p=0.001, Mann Whitney U test).

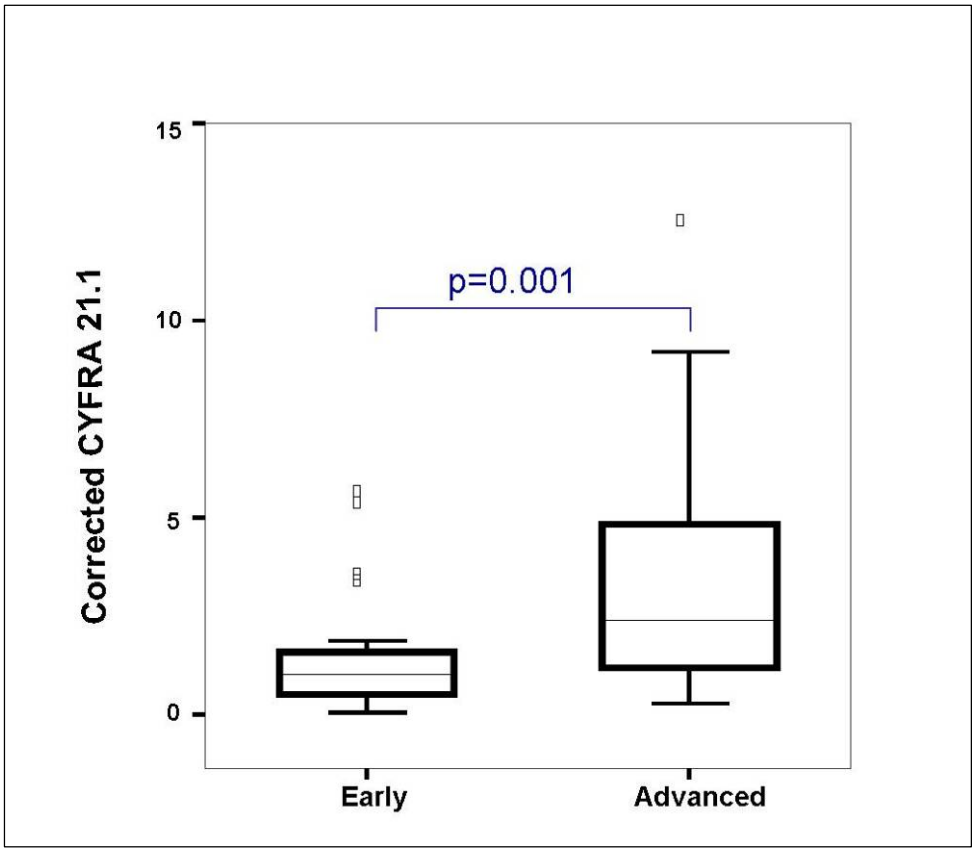


Figure 6. CYFRA 21-1 is a marker of tumour stage. Patients with advanced stage (III-IV) disease had higher CYFRA 21-1 levels than those with early stage (I-II) disease. (CYFRA 21-1 ng/ml)

Using the Kaplan-Meier method (Figure 7), an elevated CYFRA 21-1 (but not CA19-9) was an indicator of poor prognosis. For patients with levels above or below 3ng/ml, the median survival was 2 and 10 months, respectively (p=<0.001). Only 3% of patients with elevated circulating CYFRA 21-1 were

alive at 1 year and none at 2 or 3 years, compared with 42%, 22% and 10% respectively in the low circulating CYFRA 21-1 group (Table 7). The combination of CYFRA 21-1 (3.0 ng/ml) and CA 19-9 (129 IU/ml) was also a strong predictor of prognosis with median survival for those with both markers positive (+/+), one or other positive (+/-) or neither positive (-/-) of 2, 9 and 11 months respectively (p=<0.001 for both +/+ and +/- versus -/-). Lower levels of CYFRA 21-1 (1.5ng/ml) and/or CA19-9 (37 IU/ml) were not significant markers of prognosis.

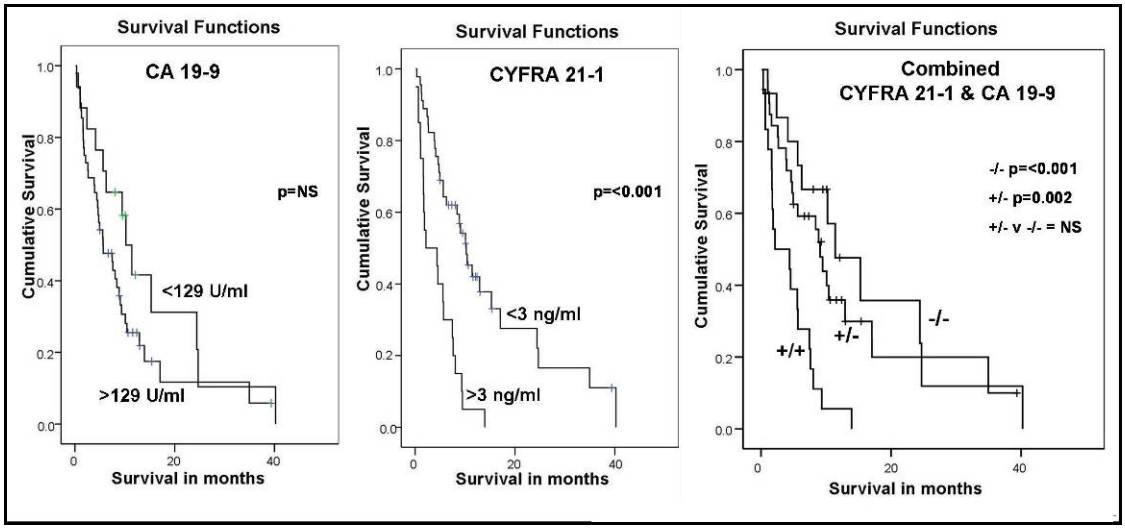


Figure 7 Kaplan-Meier survival plots showing that CYFRA 21-1 with or without the addition of CA19-9, is a prognostic indicator in patients with BTC. CA19-9 alone was not a marker of prognosis.

Table 7 Survival of patients with BTC related to circulating levels of CYFRA 21-1 and CA19-9 showing a strong association (p=<0.001 for each group) between elevated CYFRA 21-1 levels and poor prognosis.

		1 year survival	2 year survival	3 year survival	Median survival (months)
CYFRA 21-1	>3 ng/ml (n=21)	3%	0	0	2
	<3 ng/ml (n=45)	42%	22%	10%	10
CYFRA 21-1 (>3) + CA19-9 (>129)	+/+ (n=18)	5%	0%	0%	2
	+/- (n=33)	30%	16%	5%	9
	-/- (n=15)	47%	12%	10%	11

2.3.3.4 CYFRA 21-1 and CA19-9 in the surveillance of CC in PSC.

A retrospective analysis of case notes of patients with PSC at UCH, showed that in our series, all cases of CC occurred in areas of dominant strictures (DS) Figure 8. See Abstract 1 (Figure 73) in Appendix for further details (Chapman, Witmann et al. 2008; Tidswell, Chapman et al. 2009).

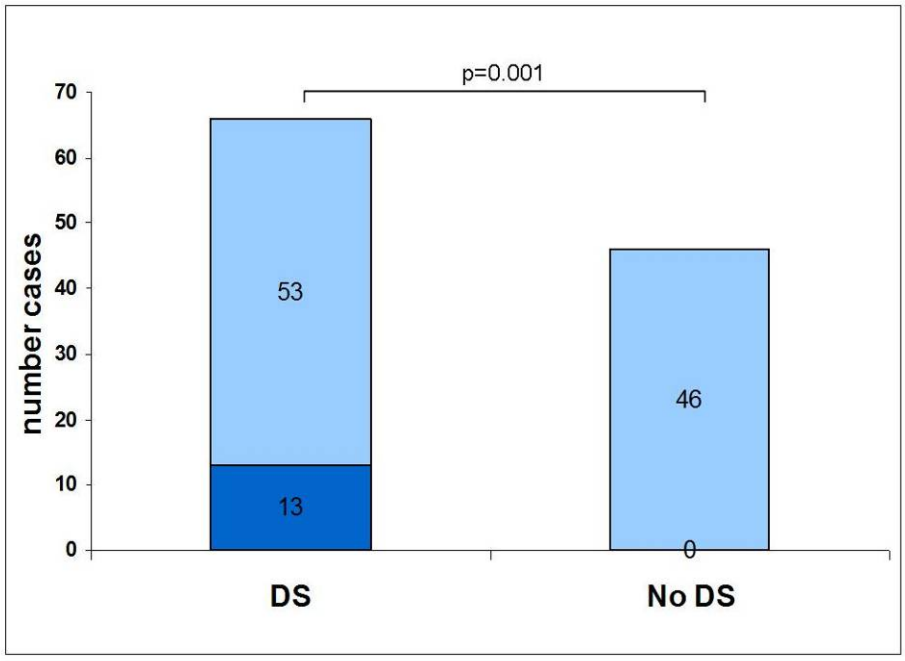


Figure 8. Number of cases of CC developing in patients with and without dominant strictures (DS) in patients with PSC at UCH. Note that CC were only diagnosed in patients shown to have dominant strictures (13/66) with none seen in those without dominant strictures.

The prognosis for patients with PSC and a dominant stricture is worse than those without this finding, (12.8 years and 19.4 years respectively) even after exclusion of those who developed CC. In those with superimposed CC, survival was a median of 5 months (Figure 9).

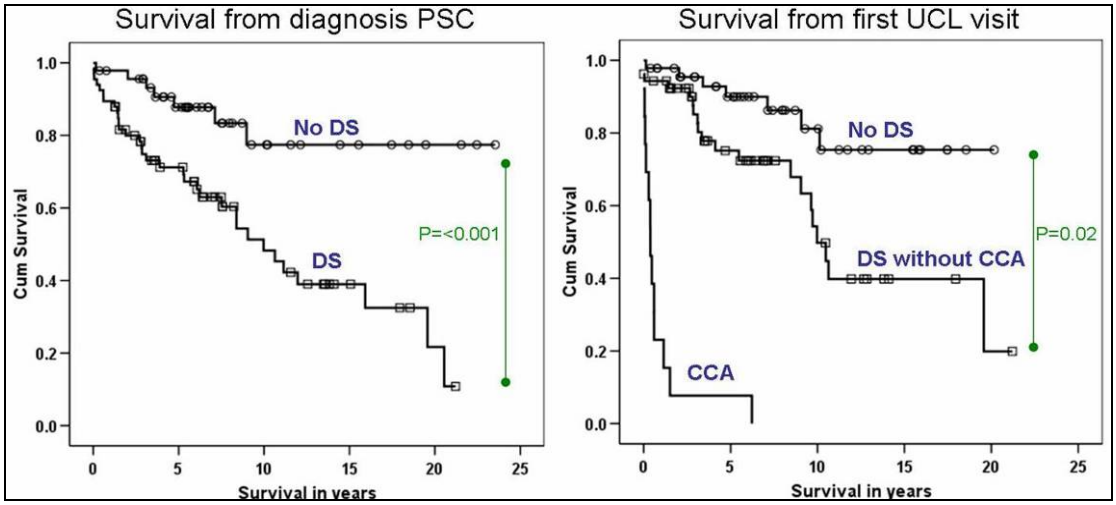


Figure 9. Kaplan-Meier survival plots for patients with and without dominant strictures. Note that patients with dominant strictures have a worse prognosis. Prognosis is particularly poor in those developing CC (right graph).

In order to address whether CYFRA 21-1 and CA 19-9 are of value in surveillance for CC in PSC, these subgroups were analysed separately to other benign diseases or non PSC related BTC. The number of cases of PSC (n=19) and PSC-related CC (n=6) were small, resulting in large standard errors. However, CYFRA 21-1 (>3.0 ng/ml) had a sensitivity of 30% and specificity of 97%. Using the combination of markers in PSC, the sensitivity and specificity were 33% and 95% for the lower cut-offs, and 16% and 100% for the higher cut-offs. In view of the small number of patients with PSC related CC, data were also assessed comparing patients with benign PSC disease against al patients with BTC (n=66) (Table 8).

Table 8 Diagnostic utility of CYFRA 21-1 and CA 19-9 for patients with BTC. Data shown are a comparison of patients with benign PSC against all patients with BTC.

	CA 19-9 >37U/ml	CA 19-9 >129U/ml	CYFRA 21.1 >1.5 ng/ml	CYFRA 21-1 >3.0 ng/ml	CYFRA >1.5 + CA19-9>37	CYFRA >3.0 + CA19-9>129
Sensitivity	79	73	56	30	51	27
Specificity	79	95	68	95	95	100
PPV	40	71	24	50	63	100
NPV	95	95	90	89	92	89

2.3.4 Discussion

Despite advances in diagnostic techniques (Moreno Luna and Gores 2006; Tischendorf, Kruger et al. 2006), BTC remains difficult to diagnose and treat with overall 5 year survival rates of 5-10% (de Groen, Gores et al. 1999). The majority of patients present with inoperable disease and have a median survival of only 6-9 months (Farley, Weaver et al. 1995; Jarnagin, Fong et al. 2001). Even those undergoing surgery with curative intent have a high rate of recurrence, in part related to late presentation, lack of reliable biomarkers and difficulties in pre-operative staging with high R1/2 resection rates (Ito, Agni et al. 2008). At present, there are no surveillance strategies proven to improve early diagnosis or outcome in high risk patients such as those with PSC. CA19-9 has a relatively good sensitivity and specificity for diagnosis of BTC but higher thresholds for diagnosis (>129 U/ml) are associated with non-resectable disease and poor surgical outcome (Levy, Lymp et al. 2005). Better non-invasive diagnostic and prognostic biomarkers are clearly needed.

In this series, which is relatively large compared with others, we have assessed the role of the tumour marker CYFRA 21-1 in patients with BTC. Our data indicate that CYFRA 21-1 is a less sensitive but more specific marker than CA 19-9 for the diagnosis of BTC (specificity 88% to 97%). The combination of CYFRA 21-1 and CA19-9 had very high specificities (96% to 100%) and positive predictive values (94% to 100%) for the diagnosis of BTC, but sensitivities were lower (27% to 45%) than for CA19-9 alone. These results are similar to reported sensitivity and specificity rates using biliary brush cytology which, despite the invasive nature of the procedure, remains

the best modality for diagnosis of BTC. Our data suggest that elevated CYFRA 21-1 or the combination of elevated CYFRA 21-1 and CA19-9 is highly suggestive of malignancy and would therefore be valuable in deciding on the need for more invasive investigations such as ERCP with biliary brushings. It was very unusual for patients with benign disease to have both markers above certain threshold values. No patient with benign disease had both markers positive using the higher cut-off levels and only two (3.4%) were positive using the lower cut-off levels. The NPV (53 to 84%) and low likelihood of malignancy with the combination of both negative results would significantly reduce the pre ERCP likelihood of identifying a malignant stricture.

In contrast to other series, our data did not support the role of CA19-9 in predicting outcome of patients with BTC. However, CYFRA 21-1 was found to be a strong predictor of prognosis with marked differences in survival in those with low or high circulating levels. This is in keeping with published data where CYFRA 21-1 predicted tumour recurrence and survival after attempted curative surgery in patients with intrahepatic cholangiocarcinoma (Uenishi, Yamazaki et al. 2008) and suggests that CYFRA 21-1 may be a powerful preoperative predictor of R0 resection and surgical outcome in patients with BTC. Also, CYFRA 21-1 may be a predictor for those undergoing surgery who may benefit from neo-adjuvant or adjuvant chemotherapy complementary to their surgical treatments.

Benign dominant biliary strictures of PSC are often difficult to differentiate from malignant strictures of cholangiocarcinoma. The only practical biomarker

for screening patients with PSC to date is CA19-9 which is reported to have a positive predictive value of 56% using a high cut-off value of 129 U/ml, and only detects advanced cases (Levy, Lymp et al. 2005). Lower cut-off levels of CA19-9 such as 40IU/ml result in only a slightly increased sensitivity with reduced specificity (57% and 84% respectively) (Charatcharoenwitthaya, Enders et al. 2008). Our CA19-9 data demonstrated an improved sensitivity (100%) but lower specificity (79%) for the diagnosis of cholangiocarcinoma in patients with PSC using a similar cut-off of 37 U/ml. However, as with many studies of patients with PSC, the number of patients included was small and insufficient to make definite conclusions on the role of CA19-9 or CYFRA 21.1 for surveillance of patients with PSC for CC, although our data do suggest that an elevated CYFRA 21-1 or CA 19-9 in patients with PSC is highly suspicious of coexistent CC (specificity 97% for CYFRA 21-1 >3 ng/ml). To have sufficient statistical power (85%), studies would require 38 patients with PSC–related cholangiocarcinoma and as such would need multi-centre collaboration.

Additional clinical tests such as CYFRA 21-1 incur additional costs. However, the cost of each ELISA plate is similar to that of CA19-9 (approximately £280) equating to less than £10 per sample, assuming 40 samples run in duplicate per plate. The CYFRA 21-1 ELISA kit is quick and easy to use and can be anaylsed in any basic laboratory with an ELISA plate reader.

The main limitation of CYFRA 21-1 and CA19-9 as tumour markers is that both have significant false negative rates. In this series there were 11% and

24% false negative rates using the combined lower and higher cut-off levels. The sensitivity of CYFRA 21-1 was too low (30% to 56%) to be useful as a screening test. In addition, the majority of our patients had advanced disease (n=42 (63%) for stage II-IV BTC) which may further lower the value of circulating CYFRA 21-1 for the diagnosis or screening for early disease and requires further investigation. It is unclear whether the presence or number of circulating tumour cells is related to tumour stage or metastatic potential. Studies using other surrogate markers of circulating tumour cells rarely identify circulating epithelial cells in healthy individuals but report the presence of such cells in patients with early localised disease (Nagrath, Sequist et al. 2007). As with other cancers, it is likely that a combination of molecular markers would improve sensitivity and specificity and this needs to be addressed with larger prospective studies (Liu, Liao et al. 2008). In addition, the use of more sensitive assays such as quantitative PCR may further improve the sensitivity of measuring circulating epithelial markers in patients with BTC or other adenocarcinomas.

One further potential counfounding factor is that patients with BTC often undergo invasive procedures such as interventional ERCP or percutaneous biliary drainage with brushing and/or stenting. Such procedures may result in transient release of epithelial cells into the circulation as a result of local trauma at sites of intervention. However, against this is that many of the blood samples were taken in the days following such procedures and the rate of false positive elevations of CYFRA 21-1 was very low in the benign groups and half that of CA 19-9 (12% and 22% using the lower cut-off levels).

With regards to the biology of BTC, the frequent finding of circulating fragments of CK19 (CYFRA 21-1) suggest that circulating tumour cells in BTC may be relatively common, a phenomenon that, as far as we are aware, has not been reported to date. It is possible that the CK19 fragments represent cellular material leaking into blood but it is more likely that they arise from CTCs. These studies support further work in the isolation and characterisation of CTC's in BTC.

In conclusion, CYFRA 21-1 alone or in combination with CA 19-9, is a highly specific biomarker for the diagnosis of BTC in patients with biliary disease. We suggest that the finding of an elevated level should prompt thorough evaluation for the presence of BTC.

Chapter 3.

3 Gene expression profiling using RNA isolated from formalin fixed, paraffin embedded tissues

3.1 Aim

To perform whole genome RNA expression profiling using surgical resection specimens (PSC, PSC-related CC and sporadic CC) in order to identify novel biomarkers of cholangiocarcinoma.

3.2 Background

Most surgical resection specimens are stored in the form of formalin fixed, paraffin embedded (FFPE) tissues. This provides a potentially useful source of valuable clinical material. The FFPE method is used clinically because it is cheap, easy to perform and produces excellent quality slides for histological assessment. Cholangiocarcinoma associated with primary sclerosing cholangitis is rare and tissue is usually only stored in the form of FFPE tissues. The main source of this tissue is the incidental or early cholangiocarcinoma in the liver explant from patients with PSC undergoing liver transplantation.

Until recently, most RNA-based molecular biology techniques have required microgram quantities of good quality intact RNA for reliable data acquisition. This has meant that only fresh or snap frozen tissue was suitable and has hampered study of the vast majority of samples stored as FFPE tissue.

Research over the last 10 years has demonstrated that successful extraction of RNA from FFPE tissue is now possible. Details of the optimal methodology for RNA isolation have been published (Lehmann and Kreipe 2001; Chung, Braunschweig et al. 2006; Coudry, Meireles et al. 2007; Penland, Keku et al. 2007; Hoshida, Villanueva et al. 2008). Furthermore, these reports have shown that, despite the high level of RNA degradation, this tissue source is often suitable for use with whole genome RNA expression profiling using microarray analysis. In order to perform such analyses, certain methodological variations are required but results are comparable to those obtained from paired fresh frozen samples (Coudry, Meireles et al. 2007).

Most earlier studies used RNA isolated from blocks or sections of whole tumour or normal liver. This potentially complicates RNA expression data because of the presence of non biliary epithelial tissue such as fibroblasts, hepatocytes and leukocytes. Some studies have used laser capture microdissection to specifically isolate the cells or tissue of interest. This technique is however also difficult because of the isolation of very small quantities of RNA, which then require RNA amplification before analysis. An example of the use of this technique is shown by Coudry et al who used laser dissected colonic epithelium from sections of surgical samples for microarray

analysis. Using paired specimens from the same source processed both by FFPE and snap freezing in liquid nitrogen, they showed that RNA isolated from both types of tissue processing resulted in similar results with a strong correlation of overall gene expression (Coudry, Meireles et al. 2007).

3.2.1 Degradation of RNA during FFPE processing

Despite obvious benefits for clinical use, FFPE processing causes significant degradation of RNA which has previously hampered further investigation. The main causes of RNA degradation in tissue samples are:

3.2.1.1 Time to fixation

As soon as tissue is resected, RNA degradation begins as a result of changes in temperature, pH, ischaemia and mostly, intrinsic RNase activity (Masuda, Ohnishi et al. 1999). It is generally acknowledged that tissues should be processed immediately as significant RNA degradation occurs within 30 minutes ex vivo. Another time-related factor is the time taken for the fixative to infiltrate the tissue. This is short with cytological or small biopsy samples but can take many hours for large resection specimens or whole organ explants. A study by Start et al demonstrated that in whole resected spleens, penetration of formalin was only 2.4mm in 24 hours (Start, Cross et al. 1992). In practice however, time of fixation for most tissues is about 1 hour for each mm of tissue and most are fixed for 24-48 hours. In general, the slower the fixation time, the poorer the RNA quantity and quality.

3.2.1.2 RNase activity

RNA is constantly broken down *in vivo* by intrinsic RNase activity which is one of the homeostatic methods for controlling RNA activity. Tissues vary considerably in the amount of intrinsic RNase and this clearly has an impact on the required speed of fixation for different tissue. Liver, biliary and pancreatic tissue in particular have very high intrinsic RNase activity and should be fixed or frozen immediately.

3.2.1.3 Formalin fixation

Formalin fixation causes modification of RNA by cross linking RNA to proteins, causing strand breakage and linkage of monomethylol groups to RNA bases. RNA degradation continues but to a much lesser extent when tissues are stored embedded in paraffin. It is therefore recommended that formalin fixed tissues should be embedded in paraffin within a few days of collection. Alternative fixatives such as ethanol reduce RNA degradation but are rarely used in clinical practice. One study of nucleic acid recovery from FFPE tissues suggests that no nucleic acid is recoverable from tissue stored in formalin for more than 1 week (Lehmann and Kreipe 2001). Of note, most large resection specimen or organ explants are stored in formalin solution.

The electrophilic attack and monomethylol addition to RNA during formalin fixation is most marked on adenine nucleotides. Therefore the polyA tail of mRNA is likely to be highly modified and this can be taken into account at later stages of RNA processing by avoiding oligo dT primers and using random hexamer primers for RNA replication and cDNA synthesis.

3.2.1.4 Heat damage during paraffin embedding

Degradation of RNA during the heating that occurs with paraffinisation almost certainly occurs but the extent of this damage is unclear and has been little studied. In contrast to the commonly held view that formalin is primarily responsible for RNA fragmentation, work done at Ambion (Austin USA) and independent laboratories, suggests that formalin induced cross linking itself does not cause RNA fragmentation but that this occurs during the heating process of paraffin embedding (Masuda, Ohnishi et al. 1999).

3.2.1.5 Degradation during long term storage

Once stored in paraffin, tissue and nucleic acids are far more stable than samples stored in formalin. The rate of RNA degradation once embedded is small but it is likely this becomes significant over many years (Ribeiro-Silva, Zhang et al. 2007). A study specifically looking at RNA extraction from new and old FFPE samples suggests that samples stored for 10 years have significantly reduced quantities and quality of RNA (Ribeiro-Silva, Zhang et al. 2007).

The remainder of this chapter is separated into two parts. Part 1 examines the use of whole sections of FFPE tissues for RNA expression in order to test methodology. Part 2 describes the use of laser capture microdissection with a view to isolating specific biliary tissues for RNA expression analysis.

3.3 Part 1: RNA expression analysis using whole sections of FFPE tissues.

3.3.1 Materials and Methods

3.3.1.1 Clinical samples and ethical approval

Clinical samples of PSC and PSC-related CC were identified via the liver transplant database at the Royal Free Hospital. Six patients with PSC-related cholangiocarcinoma were identified over a period of 15 years. FFPE samples of sporadic cholangiocarcinoma taken at surgery were available at both the UCL and Royal Free Hospital sites. Ethical approval for the research was granted by the joint UCL/UCLH research ethics committee (ref 06/Q0152/106) with additional site specific assessment and approval for the Royal Free Hospital site.

As explained later, the above samples were not retrieved for use in this project. Methodology was tested initially on other, less rare samples of colorectal, small bowel, sporadic cholangiocarcinoma and pancreatic cancer. FFPE tissue blocks were kindly provided by Ms Uzma Qureshi, Cancer Research UK Targeting and Imaging Research Group, Royal Free Campus, UCL Medical School). These samples were used to develop technique and test methodology in laser capture micro-dissection, RNA isolation and downstream analysis using quantitative real time polymerase chain reaction (qPCR).

3.3.2 RNA isolation from FFPE tissues

In view of the difficulties in extracting RNA from FFPE tissues, many researchers use commercially available RNA extraction kits designed for use with FFPE tissues. Commonly used kits include the Paradise system (Arcturus), RNeasy FFPE kit (Qiagen, Crawley, UK) and the RecoverAll kit (Ambion, Warrington, UK). Kits such as PicoPure (Arcturus, Mountain View, Ca, USA) have been developed for the extraction of picogram quantities of RNA from even single cells but these are mostly applied to non FFPE tissue samples. Similar combinations of reagents are used with similar steps in each method. Deparaffinisation is carried out using Xylene or Histoclear. Protein digestion is ordinarily performed using proteinase K or guanidine thiocyanate. RNA is precipitated and extracted using ethanol and filter columns with further clean up using DNA digestion. Depending on the digestion times, the RNA isolation process usually takes one day but can be extended up to 6 days for higher yields using prolonged deparaffinisation and protein digestion steps (Chung, Braunschweig et al. 2006). Careful RNase free technique is crucial throughout the entire process.

3.3.2.1 Methods for RNA isolation used for this work

A variety of kits and published methods for RNA isolation were used. These were initially tested using two 10μm sections of resected colon, pancreatic, small bowel and/or cholangiocarcinoma tissue for each experiment performed at least once on a minimum of 2 samples. Sections were approximately 1 year old with a tissue diameter of approximately 20mm and should therefore have

contained far more nucleic acid than would be present in small volumes of micro-dissected tissue. These larger sections were used in order to test the methodology for RNA isolation that would be used for later micro-dissected samples.

The main methods tested were:

3.3.2.2 Optimum™ FFPE RNA extraction kit (Ambion)

Sections were deparaffinised using Histoclear™ for 15mins followed by graduated ethanol washes (100% to 70% ethanol) and air dried before RNA extraction as per the manufacturer's instructions. In brief, sections were digested using a proteinase K digestion solution for 3 hours at 37^0C followed by isolation of RNA using the RNA extraction buffer, ethanol precipitation and spin column separation. Following a DNase I digestion step for 30 minutes at 37^oC, the RNA was dissolved in 10ul of RNase free water and stored at -80^oC.

3.3.2.3 Picopure™ RNA Isolation Kit (Arcturus)

Sections were deparaffinised using Histoclear™ for 15mins followed by graduated ethanol washes (100% to 70% ethanol) and air dried before RNA extraction as per the manufacturer's instructions. Sections were digested using the extraction buffer for 30 minutes at 42^oC. The resultant cell extracts were mixed with an equal volume of 70% ethanol and passed through a spin column which captures the RNA. A DNase I digestion step was performed on

the spin column before washing and elution of RNA in 11μl of RNase free water.

3.3.2.4 Modified acid-phenol, chloroform, isoamyl-alcohol method

This technique follows the methods described by Chung et al as the most effective method of isolating RNA from FFPE tissues (Chung, Braunschweig et al. 2006). Protein digestion was done using 1ml Buffer RLT (Qiagen) with 10μl B-mercaptoethanol for 48 hours at 42^{o}C. An equal volume (1ml) of acid-phenol, chloroform, isoamyl-alcohol (ratio 25:24:1) at pH 4.5 was added. The sample was mixed vigorously for 15 seconds and allowed to settle for 10 minutes before centrifuging at 13,000 rpm for 15 minutes at 4^{o}C. RNA was then precipitated using 100% isopropyl alcohol followed by centrifugation at 13,000 rpm for 8 minutes, a wash in 70% ethanol, air drying of the pellet after removal of the supernatant and re-suspension in 20μl of RNase free water for analysis and storage.

3.3.2.5 RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE Tissues (Ambion)

Deparaffinisation was performed using Histoclear™ as above. Samples were then processed as per the manufacturer's instructions. In brief, tissues were digested for 3 hours using a proteinase K based digestion buffer, nucleic acid was separated using ethanol based precipitation and capture in a spin filter column, a DNase I digestion step was performed, and lastly, RNA re-suspension and elution in 10μl RNase free water.

Variations on the above methods were performed such as using alternative deparaffinisation methods (e.g. xylene, prolonged deparaffinisation for up to three days) and variable durations of proteinase digestion steps (3 hours to 3 days). Each method was usually performed twice in order to assess its ability to isolate RNA.

3.3.2.6 Quantification of RNA recovered from FFPE sections

Studies comparing fresh, snap frozen and FFPE samples from the same source samples suggest that it is possible to recover 80% of the total RNA from frozen tissue and 30-50% from FFPE sections. The amount of RNA extracted from tissue depends on a number of factors including length of storage, volume of tissue, type of tissue, effectiveness of deparaffinisation and tissue digestion, fixative etc. mRNA is considered to be about 3% of total RNA. This has not been formally tested in FFPE extracted RNA but is assumed to be a similar fraction.

Baba et al published work on microarray analysis of laser capture microdissected (LCM) biliary epithelial cells from fresh alcohol fixed samples from patients with primary biliary cirrhosis (Baba, Kobashi et al. 2006). They demonstrated that micro-dissecting a mean of 945 cells gave a mean of 46μg total RNA after 2 rounds of RNA amplification. The starting amount of total RNA extracted was not described but assuming a 1000 fold amplification (as

per the manufacturer's information), this suggests that as little as 50pg total RNA was extracted from a mean of 945 LCM isolated cells.

Quantification of RNA can be performed using various methods but a simple, reliable and commonly used method is to measure absorbance at 260nm using a spectrophotometer. Our laboratory used the Nanodrop ND-1000 spectrophotometer (Labtech) which accurately measures RNA concentrations in small volumes of solution (1μl) at levels as low as 10ng/μl. Using this system, calculated quantities of total RNA isolated from two 10μm FFPE sections are shown in Table 9.

Table 9 Mean total RNA isolated from two 10μm FFPE sections using a selection of tested methods

Method	Mean Total RNA recovered	Spectrometric 260/280 ratio	Notes
Optimum FFPE kit	914ng	2.01	Good yield and 260/280 ratio
RecoverAll FFPE kit	360ng	1.97	Good quantity, 260/280 ratio, and plots according to Nanodrop data
PicoPure	180ng	1.47	Low yield and 260/280 ratio
Acid-phenol-chloroform	1986ng	1.86	Best yield and good Nanodrop plots using same tissues

3.3.2.7 Assessing the quality of RNA recovered from FFPE sections

The quality of mRNA extracted may be assessed using a number of methods. The most accurate method is the Agilent Bioanalyzer (see chapter 4) which was not available for use at the time this work was done. We used the following methods.

3.3.2.8 Agarose gel separation

Traditionally, the quality of mRNA is inferred by assessing the quality of total RNA which is made up primarily of 18s and 28S ribosomal RNA subunits which separate as clear bands on 1% agarose gel electrophoresis. If the bands are broken up and appear as a smear then it is assumed that the small amount of mRNA in the sample is also degraded. Samples of RNA from freshly cultured human embryonic kidney (HEK293) cells were used as a control (Figure 10). Gel separation of FFPE RNA was performed in only a few samples in view of the large quantity of RNA required for gel separation (usually >400ng for clear images).

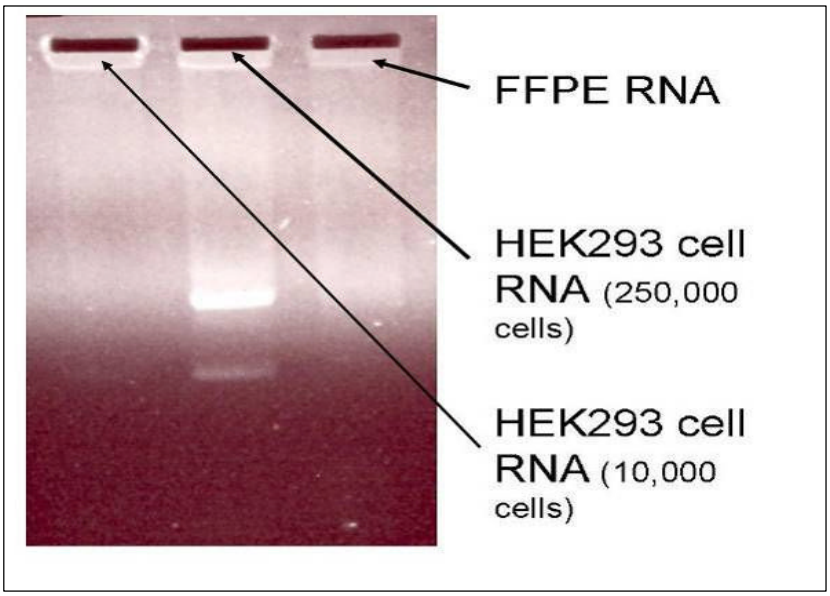


Figure 10 Agarose gel separation of RNA from freshly cultured HEK293 cells showing clear 18S and 28S bands of intact RNA and a faint smear of degraded RNA isolated from FFPE tissue. The loading quantity of RNA was smaller in the FFPE lane (150ng total RNA) than in the clear bands for 250,000 HEK293 cells seen in the central lane (250ng total RNA).

3.3.2.9 Nanodrop 260/280 ratio

A simple and commonly used assay to assess quality and purity of extracted total RNA is to measure the 260/280 absorbance ratio using a Nanodrop ND-1000 spectrophotometer. A figure of 1.80 to 2.10 is considered to be a marker of good quality RNA (see Table 9 and Figure 11).

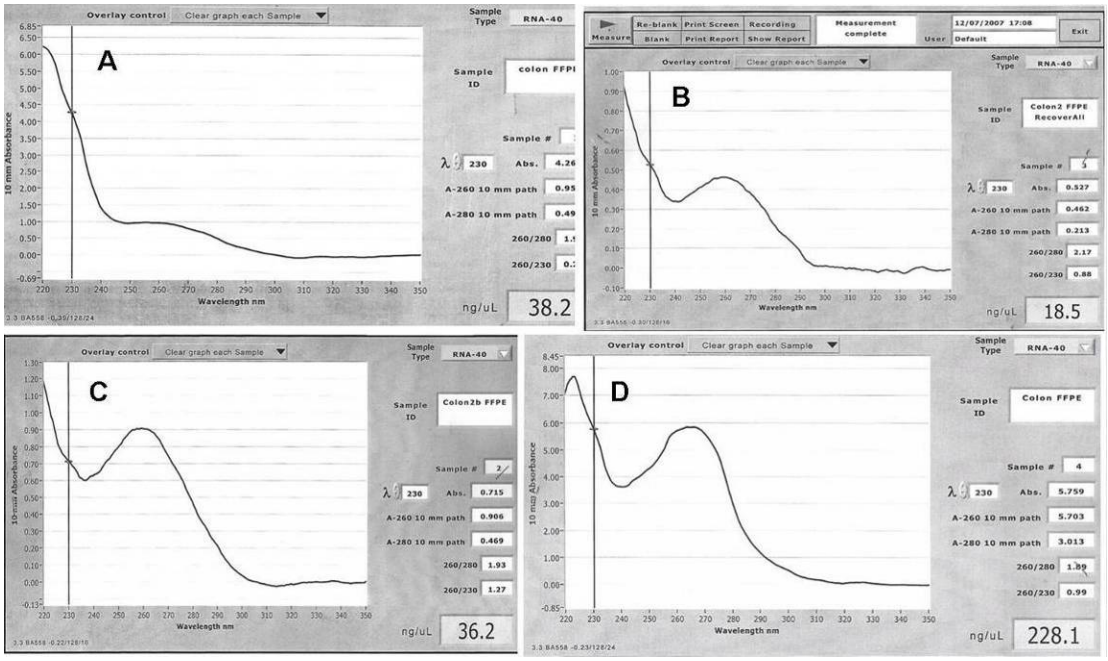


Figure 11 Examples of Nanodrop ND-1000 spectrophotometer plots of RNA peaks at 260nm. Plots show weak peak for RNA isolated using Optimum FFPE kit (A) and clear peaks of RNA at 260nm isolated from FFPE tissues using the RecoverAll kit (B &C) and the acid phenol chloroform method (D).

3.3.2.10 Functional analysis using PCR

A practical way of demonstrating the presence of satisfactory functional mRNA is to synthesize cDNA and perform RT-PCR of commonly expressed 'house keeping' genes such as β-actin and GAPDH. Because degradation results in fragmented mRNA, the cDNA synthesis step is best performed using random hexamer primers rather than poly dT primers which may result in a reduced yield because of degradation of the poly A tail on which the poly

dT primers depend. If it can be demonstrated that mRNA for reference 'house keeping' genes has been extracted than it is assumed that all or most mRNA has been extracted and is usable for downstream applications despite the RNA being degraded.

cDNA synthesis was performed using the iScript™ Select cDNA synthesis kit (BioRad). The cDNA synthesis reaction was done using 100 to 200ng of total RNA template and primed using random primers as per the manufacturer's protocol. RT-PCR reactions were performed using primer pairs for GAPDH, β-actin and 18S RNA. Primer sequences and amplicon lengths are shown in Table 10. PCR reactions were performed with a standard protocol of 5 minutes at 25^0C, 30 minutes at 42^0C and 5 minutes at 85^0C. PCR products were separated on a 2% agarose gel stained with ethidium bromide using 10μl PCR product and 1μl loading buffer (Ambion). A DNA ladder (Hyperladder II, Ambion) was used to estimate PCR product length for confirmation of appropriate amplicons. Reverse transcriptase positive (RT+) and negative (RT-) controls were used in each experiment.

Table 10 PCR primer sets used for analysis of RNA isolated from FFPE tissues

Primer	Amplicon length	Forward sequence (5' - 3')	Reverse sequence (5' - 3')
B-Actin	92	Undisclosed Eurogentec sequence	Undisclosed Eurogentec sequence
18S RNA	99	432CGGCGACGACCCATTCGAAC451	530GAATCGAACCCTGATTCCCCGTC508
GAPDH	226	108GAAGGTGAAGGTCGGAGT125	333GAAGATGGTGATGGGATTTC314
CK19	79	514TACAGCCACTACTACACGACCATCC538	592GGACAATCCTGGAGTTCTCAATG570
EGFR	68	1813TGCGTCTCTTGCCGGAAT1830	1883GGCTCACCCTCCAGAAGGTT1863
CD45	102	1353GGAAGTGCTGCAATGTGTCATT1374	1454CTTGACATGCATACTATTATCTGATGTCA1426

Further analysis was performed using real time quantitative PCR (qPCR). cDNA synthesis was done using 100 or 200ng total RNA and the iScript Select cDNA synthesis kit as above and qPCR was performed using the SYBR Green method. All reactions were carried out in 25μl reaction volumes using x2 Power SYBRGreen master mix (Applied Biosystems), 0.5μl of 10μM forward and reverse primers and 2.5μl cDNA template (equivalent to 12.5ng or 25ng input total RNA). All reactions were performed in duplicate using 96 well real time PCR plates in an Applied Biosystems 7500 Real Time PCR machine (Applied Biosystems). qPCR thermal cycling conditions were standardised as follows: 50^0C for 2 minutes, 95^0C for 10minutes followed by 40 repeat cycles of 95^0C for 15seconds and 60^0C for 1 minute.

3.3.3 Results: Part 1

Appropriate gene-specific PCR products for the house keeping genes GAPDH, 18S RNA and β-actin were identified in approximately half the samples analysed using RT-PCR. This confirmed the presence of mRNA for highly expressed house keeping genes in at least some samples (Figure 12). Interestingly, the primers for GAPDH, producing the largest PCR amplicon (or PCR product length, ie the specific nucleotide sequence synthesized by the primer set during the PCR reaction) (220bp compared to 99bp for 18S RNA and 92bp for B-actin), resulted in the weakest PCR signal (Figure 13). This highlights that optimal identification and amplification for PCR quantification was best achieved using primer sets with short PCR product lengths and that amplicon lengths of 220bp were probably too long. Similar findings using degraded RNA from FFPE have been published by others where amplification of sequences as short as 130bp resulted in weaker PCR reactions (Antonov, Goldstein et al. 2005; Coudry, Meireles et al. 2007).

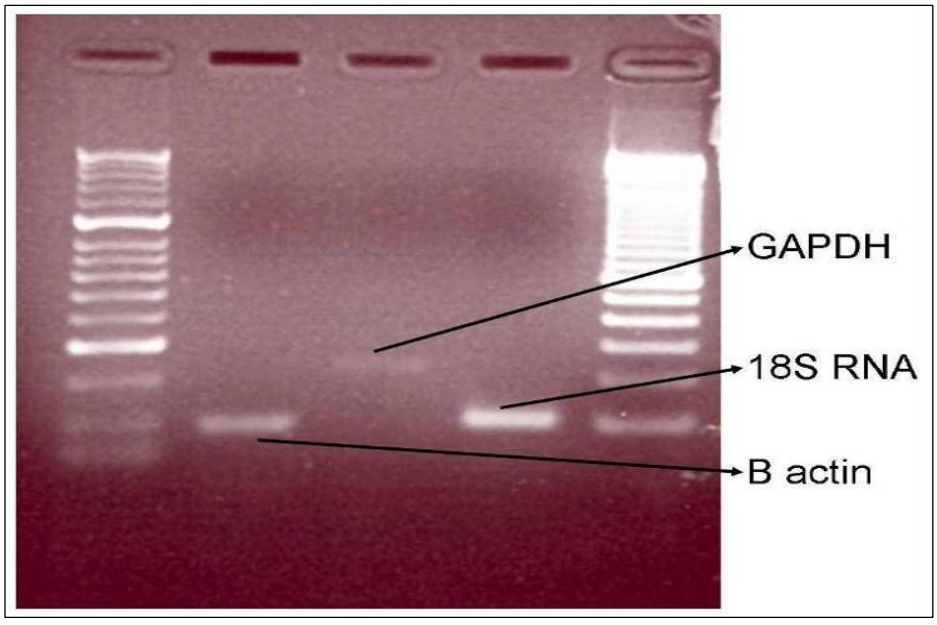
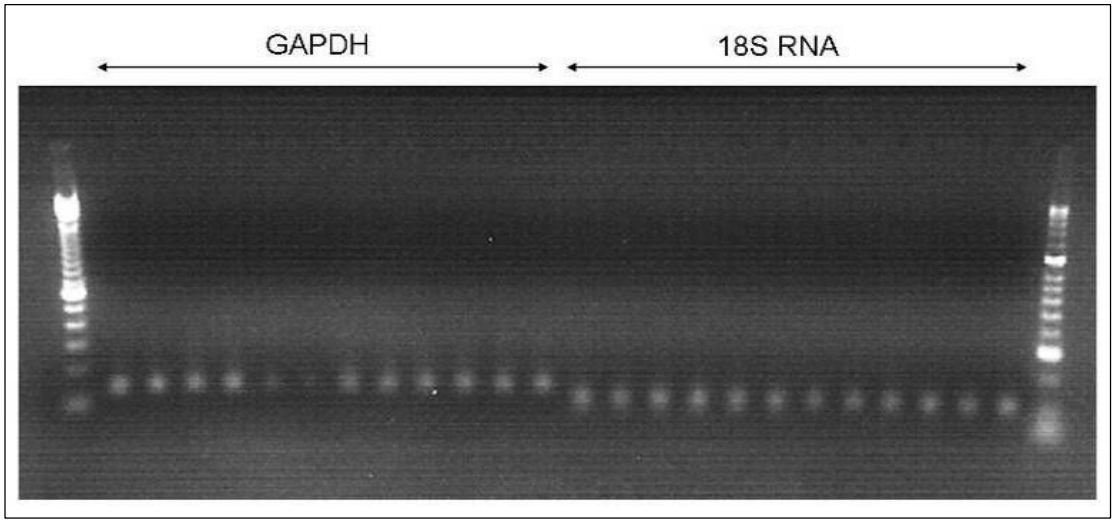


Figure 12 RT-PCR of commonly used highly expressed 'house keeping' genes using an RNA template extracted from FFPE sections. Note the presence of clear PCR products for the 3 genes tested but a weaker amplification for GAPDH which has a longer amplicon length of 220bp. DNA ladders are shown on either side of the sample lanes.

In approximately half the experiments, no PCR products were detected suggesting the absence of functional RNA from these samples. Negative results were often obtained even when relatively high quantities of RNA were measured using the Nanodrop spectrophotometer (e.g. using the acid-phenol-chloroform method). The Nanodrop machine measures absorbance at 260nm which may result in falsely elevated RNA quantification in the presence of genomic DNA contamination from incomplete DNA separation or DNase digestion steps.

Figure 13 RT-PCR using RNA isolated from 6 FFPE samples of colon and small bowel (ReceoverAll FFPE kit), showing presence of functional RNA identified using GAPDH and 18S RNA primers.



When using qPCR, most samples had threshold cycle (Ct) values of over 38 or were undetectable (Figure 14). A high Ct of over 37 represents very low levels of amplification that are considered negative. The only positive results returned Ct values of 31 to 32 for most samples, which are very low amplification thresholds for reference genes and less highly expressed genes are unlikely to be identified. However, the reverse transcriptase (RT) negative controls were also positive with similar Ct values suggesting DNA contamination in the samples (Figure 14). These RNA samples were those isolated from FFPE tissues using the modified acid phenol chloroform technique suggesting larger quantities of RNA when assessed by Nanodrop spectrophotometer. This again suggests that there was significant genomic DNA contamination during the phase separation step using this technique and results for this experiment should be considered negative. Therefore, although some experiments identified small amounts of amplified cDNA, multiple attempts failed to identify functional RNA isolated from FFPE sections.

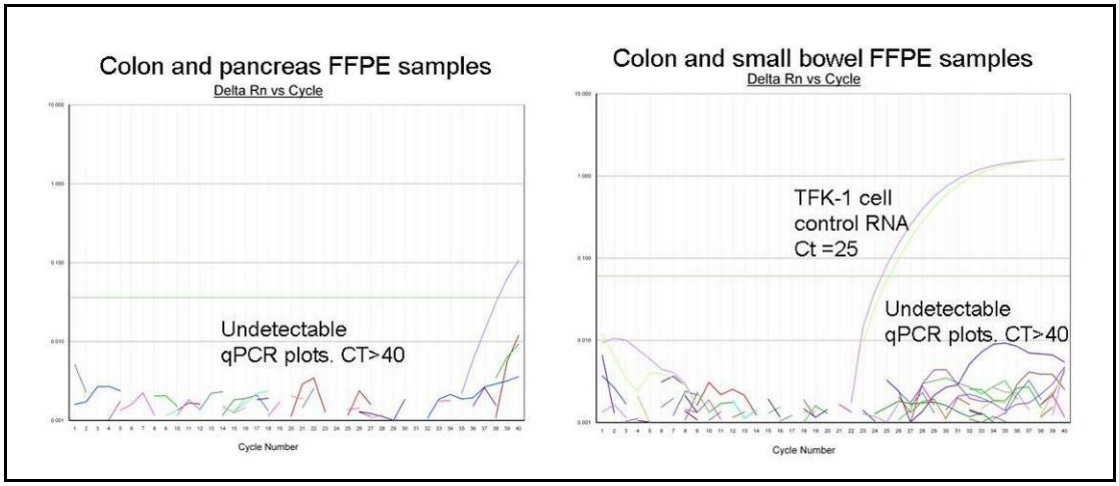


Figure 14 qPCR amplification plots for GAPDH using cDNA synthesized from RNA extracted from FFPE sections of colon, pancreas and small bowel. Note all Ct values (except one at 38) are over 40 demonstrating the absence of functional RNA. Control RNA from TFK-1 cell line amplified appropriately with a Ct value of 25. No control RNA was used in the experiment in the left plot.

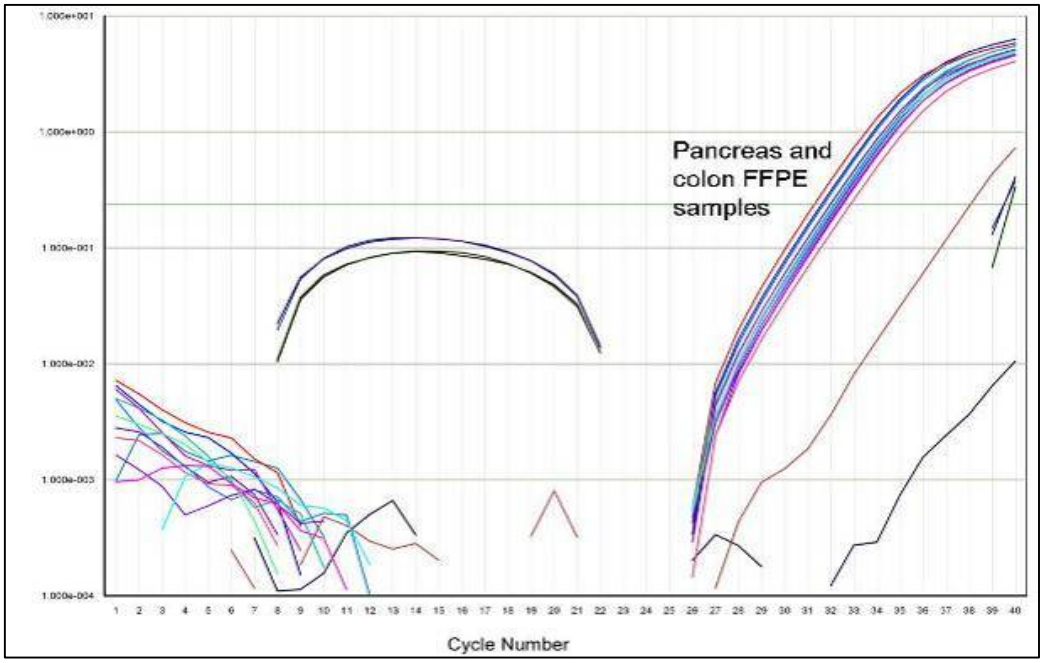


Figure 15 qPCR amplification plots for GAPDH using cDNA synthesized from RNA extracted from FFPE sections of colon and pancreas using the acid phenol chloroform technique. Ct values for the samples were mostly 31-32. However, Ct values were similar for the RT negative controls suggesting DNA contamination and the results should therefore be considered negative.

3.3.4 Conclusions- Part 1: RNA isolation using whole sections of FFPE tissues

In conclusion, RNA extracted from FFPE sections is degraded as demonstrated by smears on agarose gel electrophoresis. However, the purity as measured by Nanodrop spectrophotometer suggests protein contamination is small and RT-PCR analysis of cDNA prepared from FFPE total RNA extractions demonstrates identifiable amplified cDNA in some samples when looking at expression of highly expressed house keeping genes such as GAPDH. A number of groups have demonstrated that using the latest extraction kits, it is now possible to extract mRNA of good enough quality to assess using microarray procedures. To confirm this they have compared data using RNA extracted from fresh frozen and FFPE sections from the same patients and obtained similar results in microarray studies. These studies suggest that RNA extracted is potentially suitable for PCR in 30-50% of FFPE sections (Chung, Braunschweig et al. 2006; Coudry, Meireles et al. 2007; Penland, Keku et al. 2007). Another important principal when using RNA for downstream applications is that the fragmented mRNA is on average about 200nt long so probes and primer sets should produce amplicon lengths shorter than this. Larger amplicons of 300-400bp length are likely to give falsely negative results.

Our data using agarose gel electrophoresis and PCR of cDNA from total RNA extracted from FFPE sections of normal colon, pancreas, small bowel and cholangiocarcinoma demonstrate the extraction of useful mRNA from whole

FFPE sections. However, it is still possible that mRNA expressed at low levels may not be measurable and this cannot easily be tested without complex and expensive RNA amplification steps.

3.4 Part 2: Laser capture micro-dissection of FFPE tissues

Gene expression profiling is most appropriately performed using specific cell or tissue types. In order to isolate RNA from biliary tissue, we aimed to microdissect malignant biliary epithelium from patients with sporadic CC, PSC-related CC and benign biliary epithelium from areas distant from the tumour in the same patients.

3.4.1 General information

In order to limit further RNA degradation in the clinical samples, a clean environment with regular changes of gloves, RNase free materials and RNase free conditions using RNase digestion solutions (RNase Zap®, Ambion) were used. Training and later use of the laser capture microdissection equipment was carried out with the kind assistance of colleagues at the Cancer Research UK laboratories, Lincolns Inn Fields, London (Dr Simon Leedham).

3.4.2 Materials and Methods

3.4.2.1 Making tissue sections from FFPE blocks

The microtome was cleaned with an RNase Zap® solution and a clean blade applied. The top sections cut were kept for H&E staining as the RNA is usually markedly degraded in these sections. Sections were cut at both 5μm and 10μm thicknesses in order to test optimal size for micro-dissection, deparaffinisation and RNA yield. Sections were mounted on PALM® MembraneSlides (PALM® Microlaser Technologies, US). Automated deparaffinisation was achieved by immersing the slides in 90^{o}C xylene for 15 minutes before serial 15 minute washes at room temperature in xylene x3, 100% ethanol, 95% ethanol, 75% ethanol and lastly phosphate buffered saline (PBS) solution. Slides were air dried on a warm plate at 37^{0}C. Cut sections were used within 4 hours or stored at -80^{0}C to reduce further RNA degradation in the sections. The first cut section was stained using H&E for optimal microscopic description. Further sections were stained using Methyl green which allows basic microscopic identification of tissues for micro-dissection without causing damage to nucleic acid that is known to occur when using H&E.

3.4.2.2 Laser capture microdissection

The required tissue (initially benign and malignant colonic epithelium or pancreatic acini) was dissected out using video assisted microscopy allowing direct control of the laser dissection to the level of single cells. Approximately

3 areas of benign or malignant epithelium were dissected out from each of 5 sequential slides per patient sample and pooled for RNA isolation. Representative images are shown in Figure 16. Microdissected tissue was immediately immersed in the extraction buffer containing proteinase K (PicoPure kit) and processed as described below.

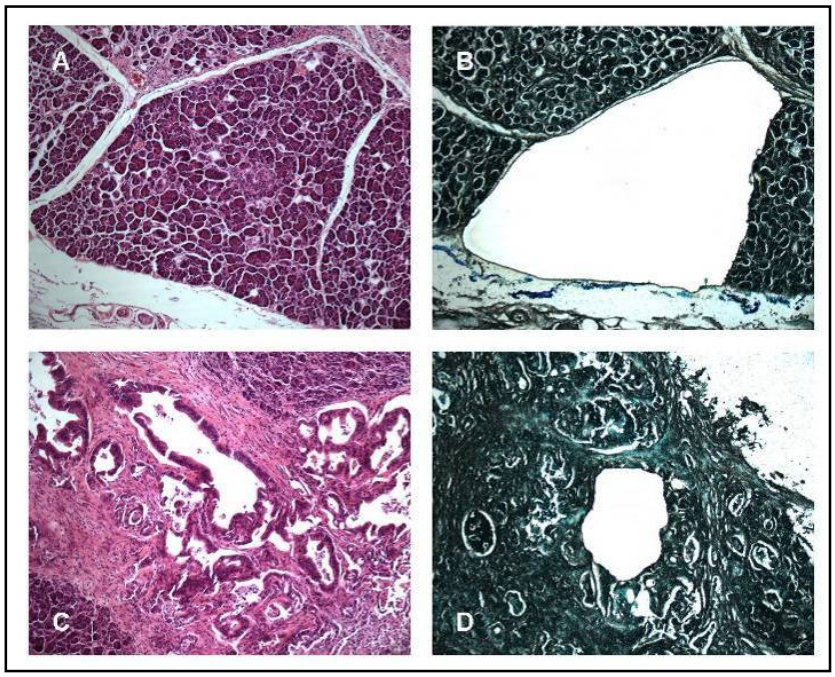


Figure 16 Representative images of laser capture microdissection of pancreatic and cholangiocarcinoma sections showing the reference slide stained with H&E (A and C) and the laser microdissected slides stained with methyl green (B and D).

3.4.2.3 RNA extraction using the PicoPure kit (Arcturus)

The RNA extraction process was performed as per the PicoPure protocol except for extending the initial step of tissue digestion in the extraction buffer. The standard protocol suggests digestion for 30mins at 42^0C before completing the extraction or storing at -80^0C for later extraction. FFPE tissues are more difficult to process and a longer proteinase digestion step helps

break down the formalin induced RNA-protein cross-linking, inhibits RNase activity, further digest tissues and opens cell membranes to improve RNA yields (Chung, Braunschweig et al. 2006). The sample was incubated in extraction buffer at 42-50^{0}C for 24hours before continuing with the RNA extraction as per the PicoPure protocol. Also of note, degraded RNA from FFPE has a reduced affinity for purification columns in the extraction process and some may have been lost during the purification process.

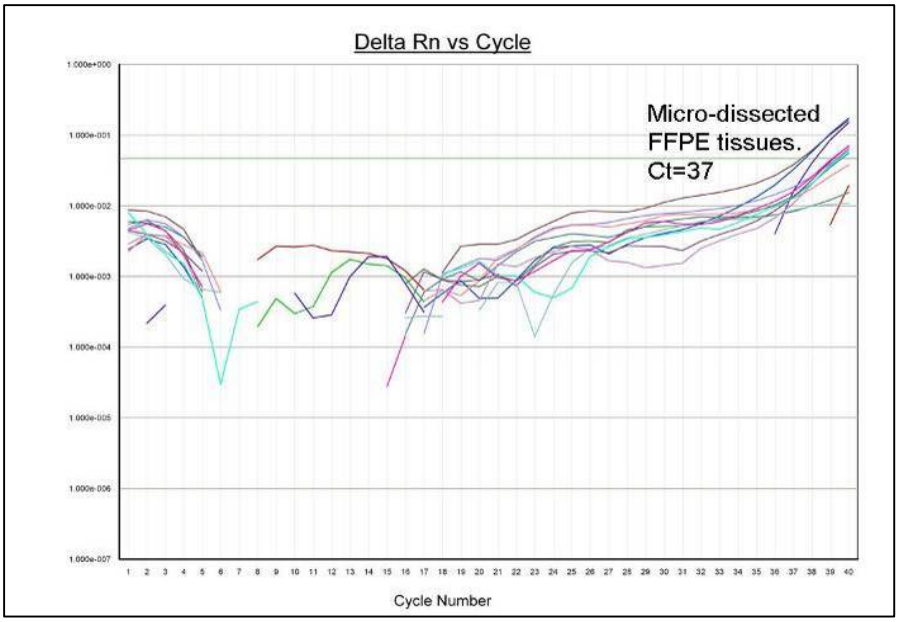
3.4.3 Results Part 2

3.4.3.1 Quantification of total RNA recovered from microdissected section of FFPE tissues

The amount of total RNA isolated from samples of LCM FFPE tissue as assessed by absorbance at 260nm by Nanodrop spectrophotometer, was 10-15ng. This is below the level of reliability for detection using these methods.

3.4.3.2 RT-PCR and qPCR using RNA isolated from microdissected pancreas and cholangiocarcinoma tissue

No identifiable bands of PCR products were seen using standard RT-PCR for the house keeping genes GAPDH, TBP or B-actin (results not shown). This suggests either very low or negative levels of functional RNA isolated using these methods. Quantitative real time PCR was used to demonstrate the presence of small quantities of functional RNA. Ct values were 37 for most of the RT+ samples and 39 or more for RT negative controls (Figure 17). These suggest absence or extremely low levels of functional mRNA.



g

Figure 17 qPCR amplification plots using RNA isolated from FFPE tissues showing weak amplification of GAPDH cDNA (Ct 37). RT negative controls were undetectable.

3.4.4 Conclusions; Part 2: Laser capture microdissection.

The modified acid phenol chloroform isoamylalcohol method appeared to be the most efficient method of isolating RNA as assessed by spectrophotometric absorbance. However, functional analysis using PCR demonstrates significant DNA contamination using this method resulting in overestimation of RNA quantities.

The results suggest that small quantities of RNA may be isolated from whole sections and LCM dissected FFPE tissues but are of insufficient quantity to be useful without increasing the total starting quantity of RNA using more tissue or an RNA amplification step. RNA amplification aims to increase the total quantity of starting mRNA, usually by a 1000 fold from the starting quantity. mRNA is relatively specifically selected for amplification by using the polyA tail of mRNA as the template to start the nucleic acid replication using oligo dT primers and T7 RNA polymerases. However, mRNA amplification methods can introduce bias as mRNA is not degraded uniformly along the 3' and 5' ends with a greater degradation at the polyA tail end, on which the amplification methods are based, being common in degraded RNA such as that isolated from FFPE tissues. Also, with each step, the aRNA (amplified RNA) product is slightly shorter than the 'parent' template so that bias can be introduced between shorter and longer mRNA sequences if later PCR primer sets used are not designed to recognise suitable areas of the aRNA product near the polyA tail end.

Sensitive assays such as qPCR for highly expressed genes failed to show useful results of gene expression and therefore genes with lower levels of expression would not be quantifiable using the types of samples available. An RNA amplification step may provide sufficient levels for quantification and qPCR but this was not attempted.

3.5 Discussion: Isolation of RNA from FFPE tissues

These and other data demonstrate that it is possible to isolate RNA from FFPE tissues and that the RNA is degraded, but useful for downstream application such as qPCR and microarray analysis. However, there are considerable limitations which are discussed in turn.

The RNA is of small quantity and highly degraded and although may be useful for further analysis with qPCR etc, primer and probe sets need to be well designed with short PCR amplicon lengths to reduce bias in results. It would seem unlikely that such RNA will be as reliable a template as RNA isolated from fresh or frozen samples and therefore these tissues will remain the sample types of choice for gene expression profiling where available.

Microdissection is a skilled and time consuming procedure and facilities are not readily available to most laboratories. However, it remains the best way of isolating specific cells or tissue types from clinical samples when investigating many aspects of cell biology and where contamination with other tissues or surrounding stroma is likely to significantly interfere with results.

The quantities of RNA isolated from micro dissected FFPE samples were too small to be quantified using standard methodologies. It may be that we were unable to recover any RNA in our samples but it is more likely that quantities were too small to be detected. Such quantities are too small for routine

protocols for whole genome RNA expression profiling using the microarray platform and require RNA amplification. Had we performed an RNA amplification step prior to cDNA synthesis and qPCR then we may have demonstrated RNA expression of genes such as GAPDH. However, the addition of an amplification step may introduce further bias which may have a major impact on gene expression profiling.

The aim of this study using the FFPE tissues was to investigate the gene expression profile of not just sporadic cholangiocarcinoma but also PSC-related cholangiocarcinoma. As discussed above, it is likely that the very old samples available for such work would have very little if any usable RNA remaining within them, even if we were to use larger whole sections of tissue which may not be representative of the malignant process. This is particularly true when large organ explant specimens, such as livers, are stored in formalin rather than paraffin embedding. We confirmed with the Department of Pathology that most of the older explant specimens of interest were stored in formalin and as such would have no significant functional RNA remaining.

Continuing with this approach to gene expression profiling for the identification of novel biological pathways or biomarkers was likely to be a very high risk approach with a high chance of failure. As such, we decided that this approach was not a feasible way of reaching the overall project goal of biomarker identification and should be abandoned in favour of using biliary brush cutology and fresh surgical tissue samples.

Chapter 4.

4 Suitability of RNA isolated from bile and biliary brushings for gene expression profiling

4.1 Introduction and background

In Chapter 3, we discussed and investigated the use of current methods in molecular biology for gene expression profiling using degraded RNA isolated from FFPE tissues. The aim of the work discussed in this chapter was to assess the suitability of RNA isolated from samples of bile and biliary brushings, for gene expression profiling.

Cytological examination of bile and biliary brushings taken at the time of endoscopic retrograde cholangiopancreatography (ERCP) have the potential to provide valuable information on the biology of benign and malignant biliary diseases. These clinical samples may be useful in developing much needed biomarkers of biliary tract cancer.

Until recently, the quantity and quality of RNA isolated from such material has been insufficient for study. As already discussed, however, advances in RNA isolation and gene expression techniques now allow small quantities of

degraded RNA, such as that extracted from formalin fixed paraffin embedded (FFPE) tissues, to be used successfully for other analytical techniques such as quantitative real time PCR (qPCR) and microarray analysis (Lehmann and Kreipe 2001; Cronin, Pho et al. 2004; Coudry, Meireles et al. 2007; Penland, Keku et al. 2007; Linton, Hey et al. 2008). RNA prepared from paired fresh frozen and FFPE tissue samples give comparable data so long as appropriate methods are used (Penland, Keku et al. 2007; Linton, Hey et al. 2008).

These technical advances allow analysis of rare and archived pathology samples as well as evaluation of degraded RNA from other sources such as bile or biliary brushings. Reports have been published describing gene expression profiling of biliary tissue using fresh frozen surgical resection material (Hansel, Rahman et al. 2003; Obama, Ura et al. 2005; Jinawath, Chamgramol et al. 2006), but there has been little published work using RNA isolated from bile or biliary brushings (Feldmann, Nattermann et al. 2006; Matull, Andreola et al. 2008). Reasons for this include the small quantity and highly degraded nature of RNA isolated from such samples although there are few published data on the quality of RNA isolated from these clinical samples.

Potential causes of RNA degradation in biliary samples include a direct effect of bile or of the x-ray contrast agents used during ERCP procedures, both of which have been shown to be cytotoxic to biliary epithelial cells in culture (Benedetti, Alvaro et al. 1997; Ju, Kim et al. 2002). A further potential confounding factor may be the presence of RNA isolated from leukocytes as a result of biliary infection, particularly in those with bile duct obstruction. There

are few published data on the quantity, the quality, or the primary source of RNA in bile, or its suitability for further investigation such as RNA expression analysis using microarray. Early work by our group suggests that there is very little RNA in bile samples and that it is degraded, but despite thise, is useful for downstream applications such as qPCR (Matull, Andreola et al. 2008). Other groups have shown that degraded RNA isolated from other biological samples, such as urine, is useful for qPCR and microarray experiments with results comparable to RNA isolated from fresh frozen tumour tissue (Mengual, Burset et al. 2006). This approach has been applied to analogous situations. Thus, whole genome RNA expression profiling was successfully applied to samples of bronchial brushings in patients with suspected lung cancer and to urinary sediment in patients with bladder cancer (Mengual, Burset et al. 2006; Spira, Beane et al. 2007).

The studies in lung cancer are particularly interesting (Spira, Beane et al. 2007). Work by this group demonstrated a field change in normal bronchial epithelium of patients with a carcinoma of the lung at a distant site. Using a panel of 80 differentially expressed genes identified by microarray (n=77 in training set and 52 in validation set using Affymetrix U133A GeneChips), patients with lung cancer could be detected from brushings of macroscopically normal bronchial mucosa with a sensitivity and specificity of 80% and 84% respectively. The addition of standard cytology of bronchial aspirates increased the sensitivity to 95% (Spira, Beane et al. 2007). This methodology may be transferable to the bile duct where the concept of a premalignant field change has been proposed in conditions such as PSC (Fleming, Boberg et al.

2001). In addition, a molecular signature of cancer or pre-cancer may be a solution to the difficult problem of the low sensitivity of standard biliary brush cytology, which is unlikely to improve using standard cytology methods.

In this chapter we investigate the suitability of bile and biliary brushings for such an approach, describe the methodology for RNA isolation and demonstrate that such samples provide useful RNA for further evaluation with downstream applications such as qPCR and microarray analysis.

4.2 Materials and methods

4.2.1 Clinical samples

Clinical samples were collected following written informed consent for ethically approved research (06/Q0152/106). Samples of bile and biliary brushings were taken at the time of clinically indicated diagnostic or therapeutic ERCP. Bile was aspirated from a major bile duct using standard biliary catheters. Biliary brushings were collected from macroscopically normal bile ducts and/or benign or malignant biliary strictures using a wire guided, sheathed endobiliary brush (Combocath Microinvasive, Boston Scientific, Notick, MA, USA), and immediately snap frozen in liquid nitrogen or lysed in TRI Reagent (Ambion,). Samples were stored at -80^{0}C until further use.

Fresh bile for cell spiking experiments was collected from free draining percutaneous biliary drains in patients with cholangiocarcinoma, or gallbladder puncture in patients having undergone planned cholecystectomy. The bile was filtered using sterile 0.45μm Millex syringe filters (Millipore, Watford, UK) in order to remove free floating cells in the donor bile. The bile was then separated into 0.5ml aliquots ready for cell spiking.

4.2.2 Leukocyte control samples

For experiments investigating the relative quantities of epithelial and leukocyte RNA, human circulating leukocytes were isolated from 20ml samples of peripheral blood (ethical approval 06/Q0152/106). Fresh samples of blood were mixed with an equal volume of RPMI culture medium. 20ml of the blood

solution was gently pipetted onto 7ml of Lymphoprep (Fresnus Kabi Norge AS, Axis Shield, Oslo, Norway) and centrifuged at 2100rpm for 20mins. The buffy coat containing the leukocytes was then removed using a pipette and spun down to a pellet and washed in PBS. The pellet was used for isolation of purified RNA as described below.

4.2.3 BTC cell lines used for positive controls

Two biliary epithelial cell lines were used for spiking experiments and controls. The first, TFK-1 (DSMZ, Germany), is a human extra-hepatic cholangiocarcinoma cell line (Saijyo, Kudo et al. 1995). The second, H69, is a human immortalised biliary epithelial cell line isolated from normal human intra-hepatic biliary epithelium (Grubman, Perrone et al. 1994). Both were cultured to confluent monolayer before use, as per published methods (Grubman, Perrone et al. 1994; Saijyo, Kudo et al. 1995).

4.2.4 Culture of biliary epithelial cells in bile and X-ray contrast agent.

In order to test whether the RNA degradation occurs *in vivo* as a result of cytotoxicity from bile or X-ray contrast agent, or ex vivo during the RNA isolation process, a series of control and cell spiking experiments were performed. A concentrated cell suspension containing 250,000 cells was spiked into 0.5ml aliquots of filtered bile or Iohexol x-ray contrast solution (Omnipaque®, GE Healthcare) at 25% and 50% concentrations diluted in

RPMI culture medium. Further aliquots of cells were transferred into an Eppendorf tube and immediately lysed in TRI reagent (Ambion) to assess the RNA integrity from cells at baseline. The spiked aliquots were incubated at 37^{o}C for 15 minutes, 30 minutes, 1, 3, 6, and 24 hours. After the designated incubation time, cells were lysed with TRI reagent and stored at -80^{o}C until RNA isolation as described bellow.

4.2.5 RNA isolation and purification

Biliary brushes were agitated in 1ml of TRI Reagent (Ambion) until most of the visible biological material had been disrupted from the brush. Total RNA was isolated using TRI Reagent according to the manufacturer's instructions. Bile appears to impair the efficiency of RNA extraction using TRI Reagent, possibly because of the volumes of bile required and possibly by chemical interaction interrupting the phase separation. We found the optimum volumes for RNA isolation are 0.5ml of bile added to 1ml of TRI Reagent. Bile also results in significant DNA contamination requiring a DNase digestion step that was performed using 4U of TurboDNase (Ambion) for 30 minutes at 37^{0}C. Total RNA was then further purified using a spin column technique (RNEasy MinElute, Qiagen) and re-dissolved in 12μl RNase free water as per the manufacturer's instructions. In addition to the brushings and bile samples, plastic biliary stents removed at the time of ERCP were also used to isolate cellular material in the early stages of testing the methodology.

4.2.6 Evaluation of RNA quantity and quality

Estimation of total RNA quantity and quality was first performed using a Nanodrop ND-1000 spectrophotometer (Labtech). 1.2ml of RNA solution was used to reliably obtain readings from the spectrophotometer, leaving 10.8ml of RNA solution remaining for further analysis. Readings were taken pre and post DNase digestion and cleanup steps but were unreliable pre DNase digestion because of the large quantity of contaminating DNA. Data recorded included estimation of RNA concentration (ng/ml) by measuring absorbance at 260nm, and 260/280 absorbance ratios as a measure of RNA quality and purity. 260/280 ratios of 1.8 to 2.1 were considered to be markers of relatively good quality and purity of RNA.

An Agilent 2100 Bioanalyzer was used after RNA clean up to further accurately assess RNA quantity and quality. The Agilent RNA integrity number (RIN) and electropherogram plots were used to document RNA quality and calculate fragment size. The larger quantities of RNA isolated from biliary epithelial cells cultured in bile or x-ray contrast was also assessed by gel separation on a 1% Agarose gel before and after DNase digestion and cleanup.

4.2.7 cDNA synthesis

Purified total RNA was reverse transcribed using the iScript Select cDNA synthesis kit (Bio-Rad) as per the manufacturer's protocol. Reactions were

performed in 20μl volumes using 150ng or 200ng total RNA and primed with random hexamers in view of the fragmented nature of the RNA.

4.2.8 Quantitative real time PCR

PCR reactions were carried out in 25μl volumes using the qPCR Master mix plus dNTP kit (Eurogentec) with 2μl of cDNA (equivalent to 15ng or 20ng RNA) sample template per reaction. PCR reactions were performed in duplicate using the SYBR Green detection method and an Applied Biosystems 7500 Real Time PCR system (Applied Biosystems). Thermal cycling conditions were set at 50^{o}C for 2 minutes, 95^{o}C for 10 minutes followed by 40 repeats of 95^{o}C for 15 seconds and 60^{o}C for 1 minute. Quantitative values for gene expression were calculated using the ΔCt method normalizing to GAPDH (NM002046.3) as the reference gene. qPCR measurement of GAPDH expression was also assessed using 4 different primer sets producing PCR products of increasing lengths (87nt, 131nt, 220nt and 405nt lengths).

When using clinical samples of bile and biliary brushings, expression of CK19 (NM 002276.3), EGFR (NM 005228.3) (markers for biliary epithelial cells) and CD3 (NM 000733.2) and CD45 (NM 080922.1) (leukocyte markers), were measured in order to ascertain the primary origin of the RNA. Expression of MUC4 (AF058803) and MUC5AC (Z48314) were also measured in clinical samples. Primer pairs used are listed in Table 11.

Table 11 Primer sets used for qPCR experiments

Primer	Amplicon length (bp)	Forward sequence (5' - 3')	Reverse sequence (5' - 3')
GAPDH 87	87	556TGCACCACCAACTGCTTAGC575	642GGCATGGACTGTGGTCATGAG621
GAPDH 131	131	156CACCAGGGCTGCTTTTAACTCTGGTA181	286CCTTGACGGTGCCATGGAATTTGC263
GAPDH 226	226	108GAAGGTGAAGGTCGGAGT125	333GAAGATGGTGATGGGATTTC314
GAPDH 406	406	625ATGACCACAGTCCATGCCATCA 646	1030GCTTGACAAAGTGGTCGTTGAG1009
CK19	79	514TACAGCCACTACTACACGACCATCC538	592GGACAATCCTGGAGTTCTCAATG570
EGFR	68	1813TGCGTCTCTTGCCGGAAT1830	1883GGCTCACCCTCCAGAAGGTT1863
CD3	88	161GGCAAGATGGTAATGAAGAAATGG184	249AGGGCATGTCAATATTACTGTGGTT225
CD45	102	1353GGAAGTGCTGCAATGTGTCATT1374	1454CTTGACATGCATACTATTATCTGATGTCA1426
MUC4	101	1499GCCCAAGCTACAGTGTGACTCA1520	1600ATGGTGCCGTTGTAATTTGTTGT1578
MUC5AC	102	273TCCACCATATACCGCCACAGA293	375TGGACGGACAGTCACTGTCAAC354

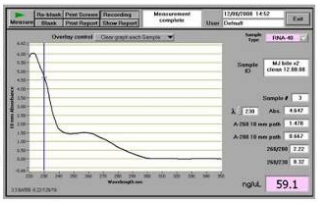
4.3 Results

4.3.1 Quantity of RNA recovered from bile and biliary brushings

The quantities of purified total RNA isolated from the clinical samples as assessed by Nanodrop ND-1000 spectrophotometer were as follows;

- Bile (n=26);
    - median total RNA 148 ng (range 0 – 535 ng).
    - 260/280 ratio 2.17 (1.34 - 3.25).
- Biliary brushings (n=51);
    - median RNA 759ng (range 44 - 2640ng).
    - 260/280 ratio 2.04 (1.2 to 2.74).

Sample ID	User ID	Date	Time	ng/ul	A260	A280	260/280
EP bile RNA clean 31.03.08	Default	31/03/2008	15:15	48.65	1.216	0.561	2.17
VB bile RNA clean 31.03.08	Default	31/03/2008	15:17	18.91	0.473	0.323	1.47
MT bile RNA clean 31.03.08	Default	31/03/2008	15:18	8.90	0.222	0.089	2.49
HE bile RNA clean 31.03.08	Default	31/03/2008	15:18	10.37	0.259	0.123	2.10
NC bile RNA clean 31.03.08	Default	31/03/2008	15:19	47.65	1.191	0.658	1.81
JP bile RNA clean 31.03.08	Default	31/03/2008	15:20	15.74	0.393	0.154	2.56



Sample ID	User ID	Date	Time	ng/ul	A260	A280	260/280
MN brush RNA clean 17.10.07	Default	17/10/2007	15:00	167.77	4.194	1.927	2.18
JJ brush RNA clean 17.10.07	Default	17/10/2007	15:01	161.88	4.047	1.885	2.15
JH brush RNA clean 17.10.07	Default	17/10/2007	15:02	229.59	5.740	2.710	2.12
MN stent RNA clean 17.10.07	Default	17/10/2007	15:03	733.65	18.341	9.043	2.03
JH stent RNA clean 17.10.07	Default	17/10/2007	15:04	558.79	13.970	7.017	1.99
KW stent RNA clean 17.10.07	Default	17/10/2007	15:05	298.22	7.456	3.765	1.98

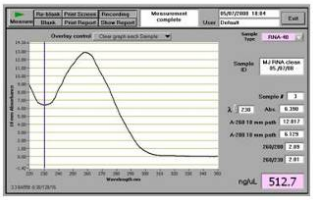


Figure 18. Representative examples of Nanodrop ND-1000 spectrophotometer data of RNA quantity and quality as assessed by 260/280 nm wavelength absorbtion ratios. RNA quantity and quality are higher in the biliary brush samples than bile samples.

4.3.2 Quality of RNA recovered from bile and biliary brushings.

RNA integrity as assessed by Agarose gel electrophoresis was mostly investigated where larger quantities of RNA were available in the cell culture spiking experiments in order not to lose valuable RNA from bile and biliary brushings. In the few clinical samples assessed by gel electrophoresis, faint smears of RNA were seen, suggesting the RNA to be highly degraded (Figure 19). Agilent Bioanalyzer plots of purified RNA from bile and biliary brushings demonstrated highly degraded RNA with low RIN scores and short fragments of RNA estimated to be primarily 100 to 600nt long (Figure 20 and Figure 21). Median RIN scores were 2.4 (range 1 to 3.9) for bile and 2.4 (range 1 to 5.8) for biliary brushings.

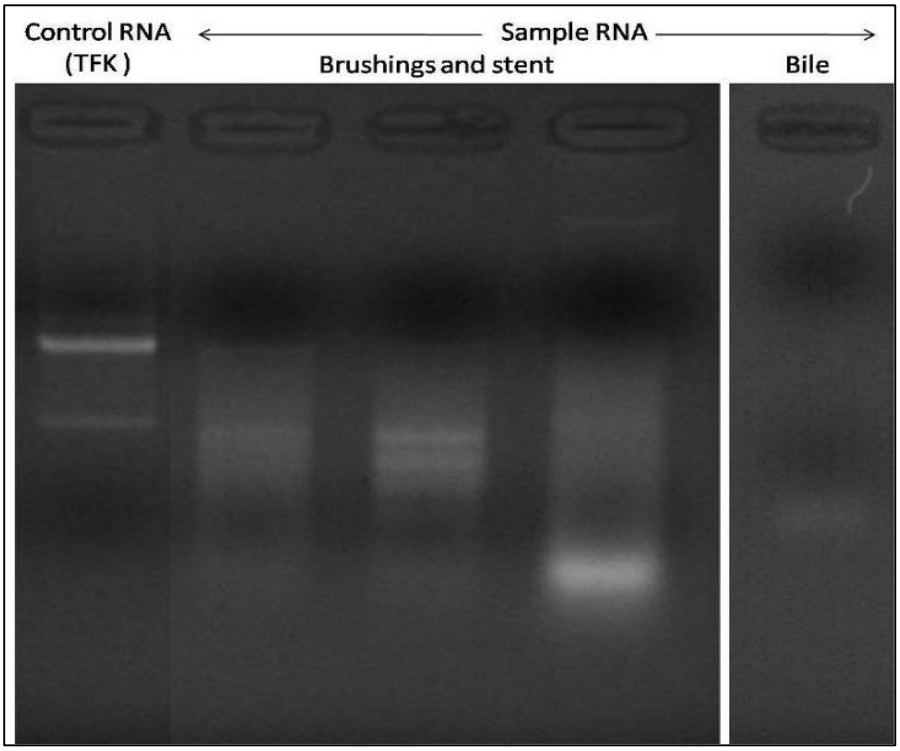


Figure 19. Agarose gel separation of RNA isolated from control TFK-1 cells (intact bands), samples of biliary brushings and biliary stent (degraded) and faint smear from bile (degraded).

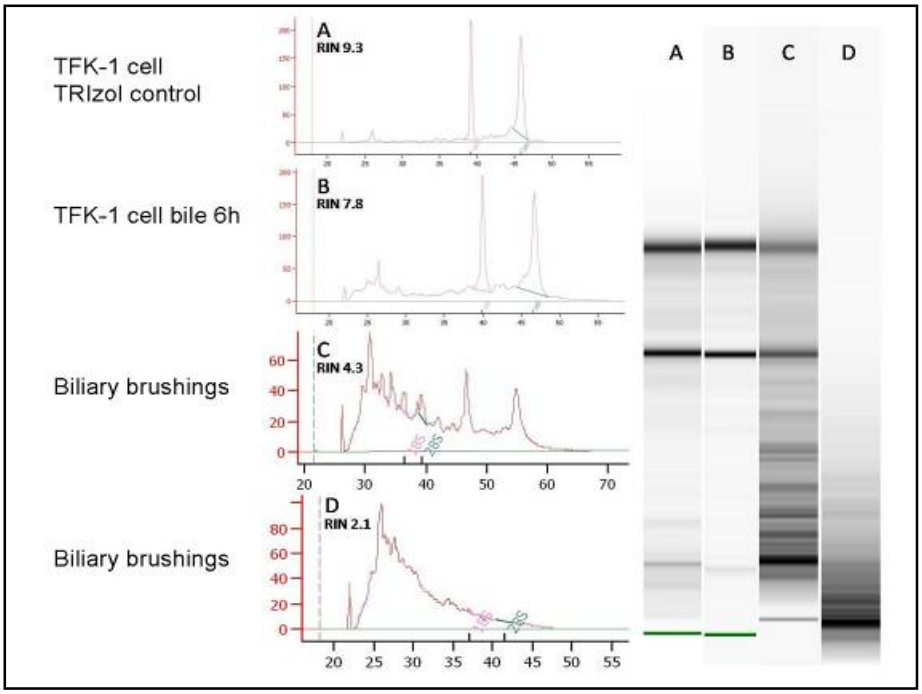


Figure 20. Representative Agilent Bioanalyzer plots of RNA isolated from A) control TFK cells [intact], B) TFK cells incubated in bile (6 hours) [minimal degradation], C) biliary brushings, [partly degraded], and D) bile [highly degraded].

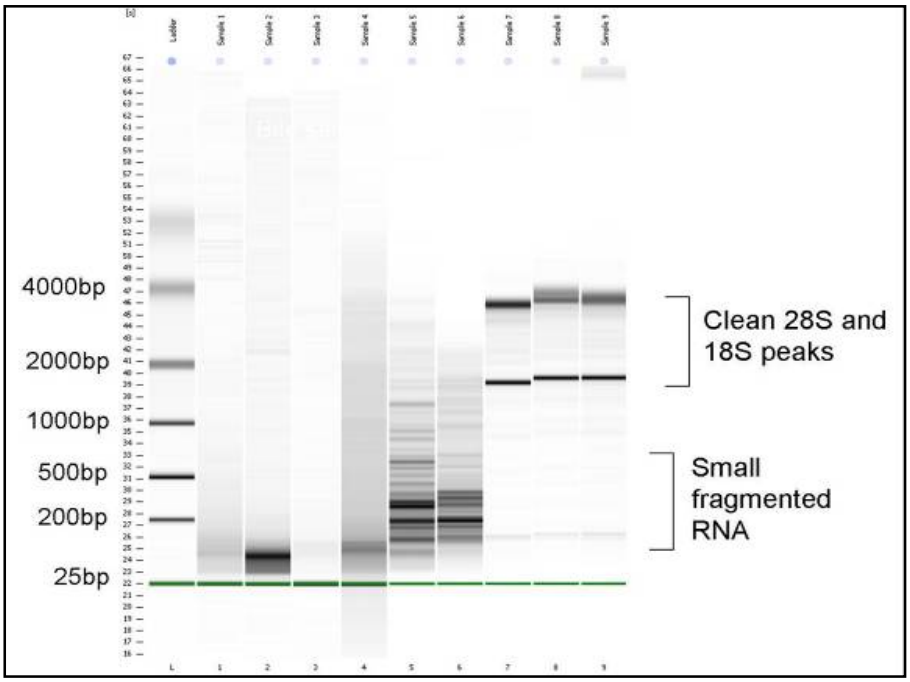


Figure 21. Examples of Agilent Bioanalyzer electropherograms showing small size (100-500 nt) of the degraded RNA isolated from clinical samples of bile and biliary brushings (lanes 2-7). The three electropherograms on the right show clear peaks of 18S and 28S RNA demonstrating intact RNA from controls samples (TFK-1 cells)

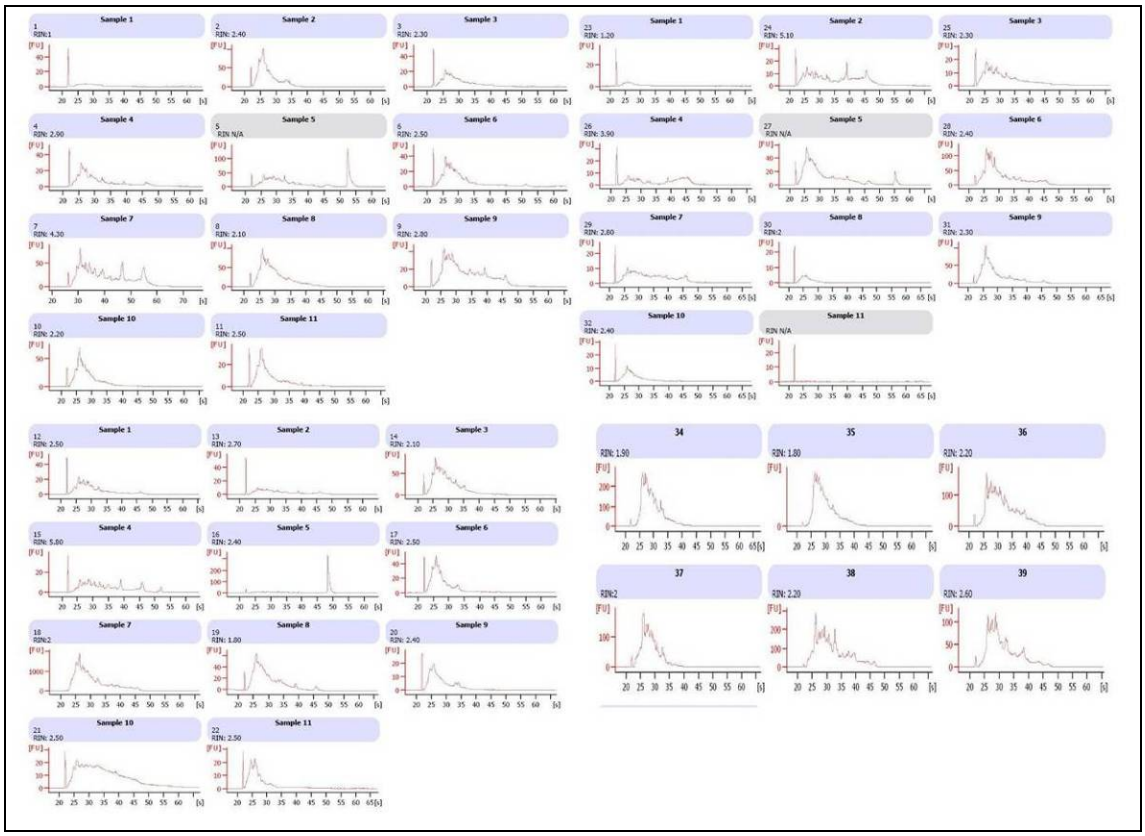


Figure 22 Agilent Bioanalyzer plots of RNA isolated from biliary brush samples (n=38) showing moderately (identifiable 18S and 28S peaks and RIN 3-6) and severely (no identifiable 18S and 28S peaks and RIN <3) degraded total RNA.

4.3.3 Quality of RNA isolated from PBMC's and BTC/biliary epithelial cells (BEC) cultured in bile or x-ray contrast agent

Purified RNA isolated from freshly cultured PBMC's and BTC/BECs was high quality intact RNA as assessed by both agarose gel electrophoresis and Agilent Bioanalyzer (RIN scores >8, Figure 23 and Figure 24). This demonstrates that the methodology for RNA isolation is suitable and not responsible for the RNA degradation found in the clinical samples. Although the quantity of RNA isolated from BECs cultured in whole bile was reduced,

the quality of RNA extracted was excellent, even after 24hours culture in bile (RIN 7.8 to 9.0). Similar findings were seen in BECs cultured in 25% and 50% x-ray contrast (RIN 7.8 to 8.9) but with less marked effect on RNA quantity, which was thought to be related to an effect of bile on the phase separation using TRI Reagent. These data suggest that short term exposure to bile or x-ray contrast agents are not responsible for the RNA degradation found in clinical samples of bile or biliary brushings.

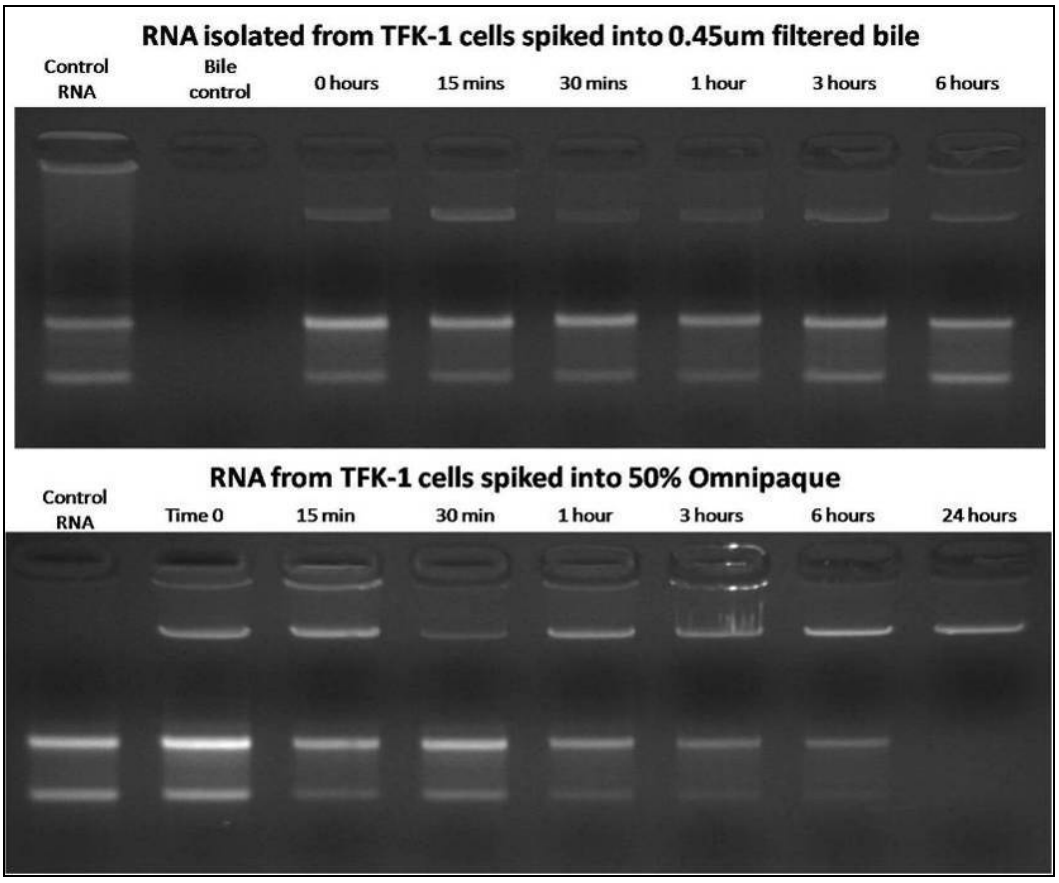


Figure 23. Agarose gel separation of RNA isolated from TFK-1 cells cultured in filtered bile (top) and 50% Omnipaque (bottom) showing clear bands of intact RNA up to at least 6 hours. Note the bands of DNA contamination seen near the starting wells.

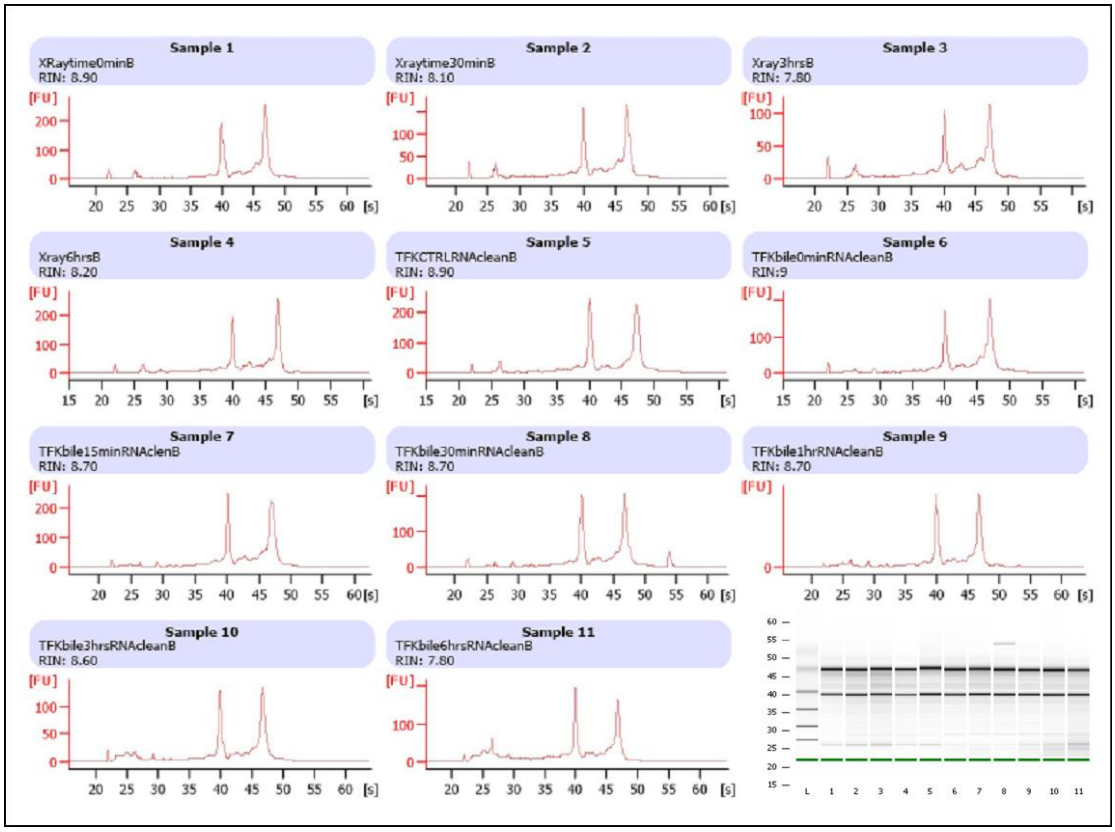


Figure 24. Representative Agilent Bioanalyzer electropherograms for purified RNA isolated from BTC cells cultured in bile and x-ray contrast agent. All samples were of good quality, similar to the control samples with clear rRNA peaks and RIN scores mostly >8, even when incubated in bile or x-ray contrast for up to 6 hours. The same samples are shown as electrophoresis strips showing clear 18S and 28S bands of intact RNA (bottom right).

4.3.4 qPCR analysis of degraded biliary RNA

To test whether the RNA isolated from clinical samples was reliable for gene expression analysis by qPCR and to test the effect of PCR amplicon length, GAPDH mRNA expression was assessed using 4 different sets of primers. These were designed to amplify regions of the GAPDH transcript ranging from 87 to 406 base pairs in length. In the clinical samples the lowest Ct values (*i.e*. greatest GAPDH mRNA expression) were consistently obtained when using primers amplifying the shortest region (87bp) of the GAPDH transcript.

However, the Ct values from the same samples rose significantly, in a stepwise fashion, when primers generating larger amplicons were used (Figure 25). This effect was primarily seen in highly degraded RNA from bile with much less significant variation using partially degraded RNA for which primer amplicons of up to 226bp length appear to be satisfactory. In contrast, no shift in the Ct values was observed when assessing GAPDH mRNA expression in control TFK-1 or H69 cells, regardless of the primer set used. As described by others (Paska, Bogi et al. 2004), primers with larger amplicons (e.g. 406bp), result in a significant rise in amplification thresholds and are not suitable for qPCR using the SYBR Green method (Figure 25). When assessing other genes of interest (CK19, EGFR, MUC4, MUC5AC and CD45) in clinical samples (Figure 25), primers amplifying regions ≤100bp and the ΔΔCt method can be used to assess relative gene expression (Livak and Schmittgen 2001).

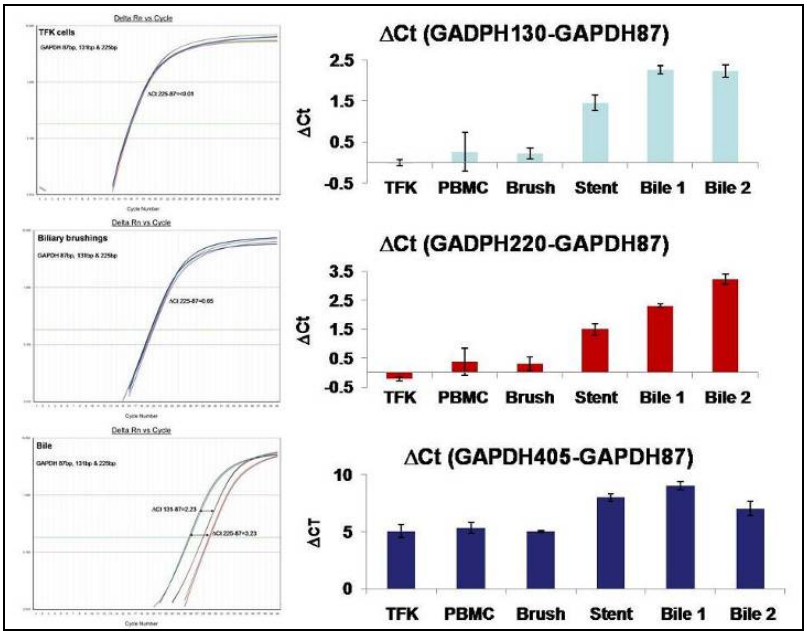


Figure 25. qPCR amplification plots (left) and ΔCt calculations of gene expression using GAPDH primer pairs with increasing amplicon lengths. Note the separation in amplification plots and larger ΔCt when using PCR primers sets with long amplicon length with degraded RNA from bile, but not with intact control (TFK-1) RNA.

4.3.5 Relative expression of epithelial and leukocyte RNA in biliary brushings

Expression of CK19 and CD45 were measured as markers of epithelial and leukocyte RNA respectively. TFK-1 cells (BTC cell line) and human PBMCs were used as positive controls to assess the level of expression in pure samples. Note that the Ct values were very high and close to GAPDH for genes of interest in the control samples but very low (over 35) or negative in the other cell type (Figure 26). For clinical samples, the ΔCt values compared to the reference genes were used an estimate of the relative proportion of the two types of RNA in the samples.

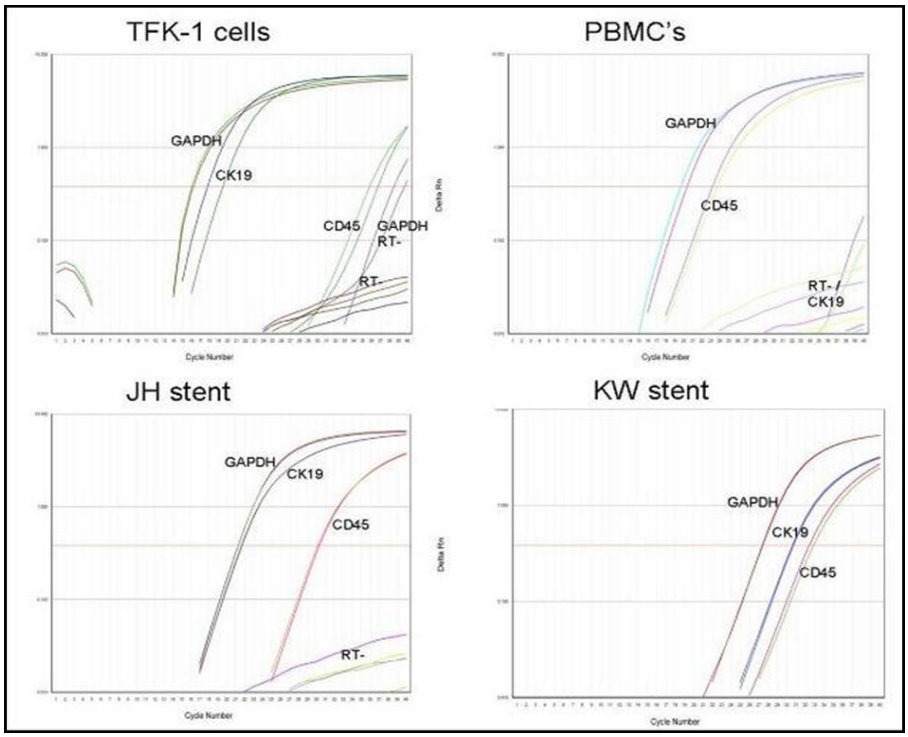


Figure 26. Confirmation of suitability of qPCR primer sets. Note that the Ct value for CK19 in TFK-1 (BTC) cells and CD45 in PBMCs respectively are very low, demonstrating high expression levels near to those of the house keeping gene GAPDH. Conversely, levels of CD45 are very low (CT >35) and negative for CK19 in PBMCs confirming specificity of the primer sets. Relative expression of CK19 was higher than CD45 in stent samples but there is significant variation in both.

The relative expression of epithelial RNA was consistently higher than that for leukocyte RNA suggesting that the RNA isolated from bile or biliary brushings is primarily of epithelial origin. This pattern of gene expression can be seen on individual sample gene plots (Figure 27) and when grouped for calculation of mean figures (Figure 28). Also of importance was that there was no significant difference in relative expression of CK19 in benign versus malignant samples (1.02, CI 0.98-1.06) suggesting that both have similar quantities of epithelial RNA and should therefore be suitable for comparison using the $\Delta\Delta$Ct method. Relative expression of CD45 (1.65, CI 1.48-1.83) was slightly higher in malignant samples raising the possibility of bias in genes expressed by leukocytes. However, as shown above, the relative quantities of leukocyte RNA are much smaller (difference in ΔCt approximately 6, equating to 64 fold higher epithelial RNA assuming that the respective markers of cell type are equally highly expressed by the cells of interest) suggesting that a slight difference in total leukocyte RNA is unlikely to significantly impact on results in genes of interest.

Although not an accurate method for assessing relative quantities, one may estimate this by comparing the DCt for each against the reference gene (18S). This is not accurate as it assumes that i) 18S is equally expressed by both cell types (likely true), ii) 1Ct accurately represents x2 fold change (likely true) and iii) both CD45 and CK19 are equally highly expressed in their respective cells (assumed but not quite true despite small ΔCt for each showing both expressed at levels near to GAPDH in their respective cell

type). However, based on these assumptions and using the ΔCt for CD45 (21) and CK19 (14.8), in biliary brushings, there was approximately a 77 fold greater ratio of epithelial to leukocyte RNA.

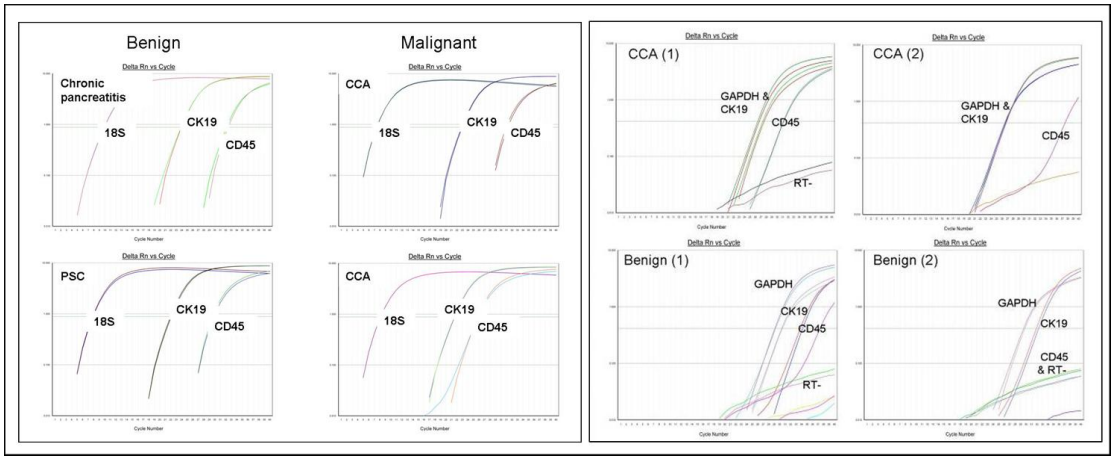


Figure 27. qPCR amplification plots for RNA isolated from biliary brushings showing a higher relative expression of the epithelial cell marker (CK19) than leukocyte marker (CD45). A similar pattern of epithelial and leukocyte RNA is seen in malignant and benign disease. The two boxes represent data from two different experiments, one using GAPDH (left) and the other 18S (right) as the reference genes.

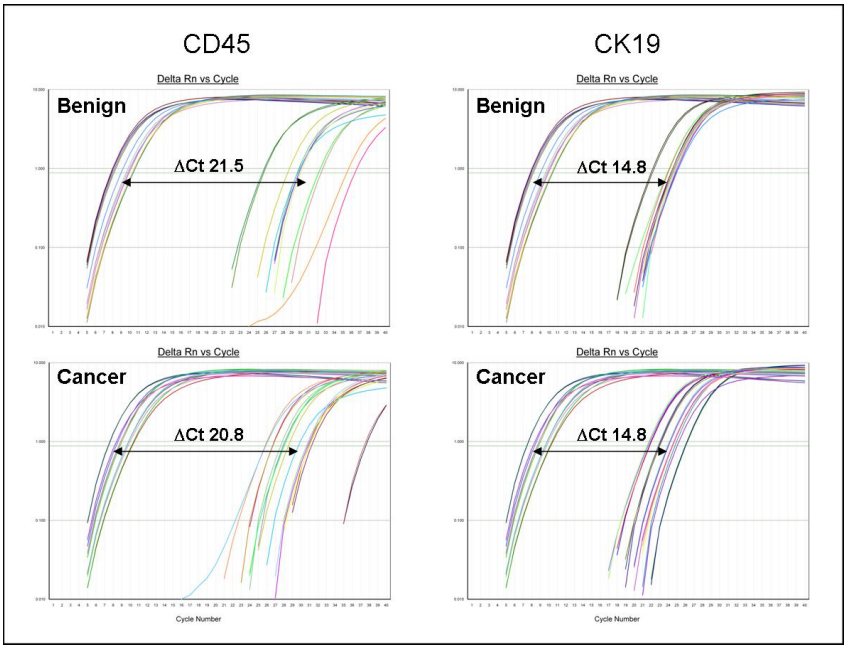


Figure 28. Amplification plots for CD45 and CK19 showing mean ΔCt calculated against the 18S reference gene. In both benign and malignant samples, expression of CK19 is similarly higher than that for CD45. Also the expression of CD45 varies considerably between samples.

4.3.6 Real time qPCR using samples of bile and biliary stents

Similar qPCR results were obtained using highly degraded RNA isolated from bile. Gene expression was quantified for 18S, GAPDH, CK19 and CD45 using the ΔCt method (Figure 30 and Figure 31).

The input quantity of RNA is calculated to be similar for all qPCR reactions and so the expression of the reference gene (GAPDH or 18S) should be similar. However, unlike the biliary brush samples, there was a large variation in the Ct values for the reference genes when using samples of bile (Figure 29). This may be partly due to errors in spectrophotometric assessment of calculations at very low concentrations and/or due to inefficient amplification of PCR products with the highly degraded RNA in bile or genomic DNA contamination.

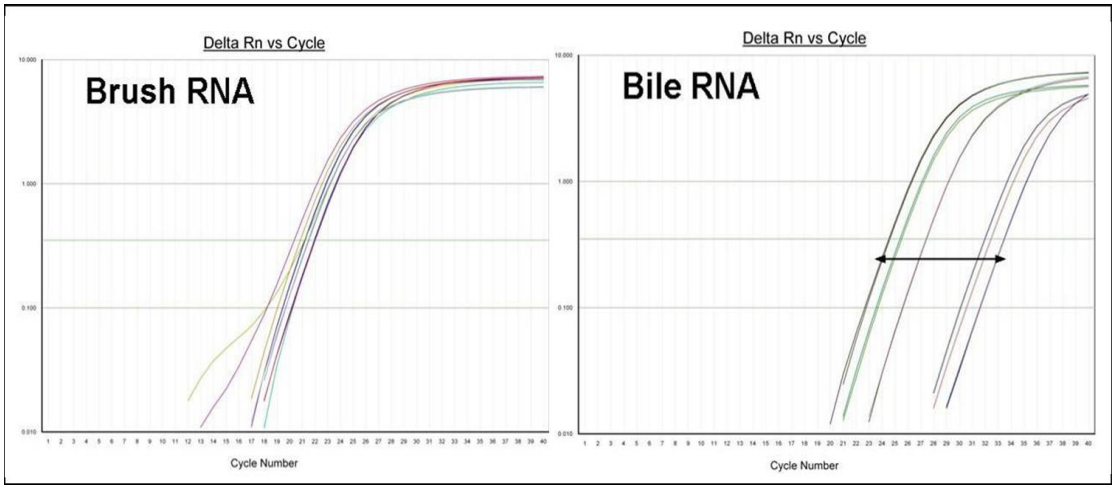


Figure 29. Despite equivalent starting quantity of RNA as measured by Nanodrop spectrophotometer (150ng per cDNA synthesis) and equivalent calculated cDNA loading (1μl or 7.5ng equivalent RNA), there is considerable variation in Ct values for GAPDH in the bile samples (arrow) but not in the brush samples. This suggests that the RNA isolated from bile is either poorly quantifiable or the higher level of RNA degradation in bile results in greater variability in gene expression, even when using primer pairs with short amplicons (87bp for GAPDH). Alternatively, genomic DNA contamination may result in inaccurate estimation of the starting concentration of RNA. However, the ΔΔCt method may still be useful on the assumption that the effect of RNA degradation is similar for other measured genes within a sample.

When using clinical samples of bile (n=12) and 18S as a reference gene, the relative expression of CK19 mRNA was much higher (ΔCt 18.15) than that of CD45 (ΔCt 24.68) showing a significantly higher contribution of epithelial RNA than leukocyte RNA in bile (Figure 31). When comparing the relative expression of epithelial and leukocyte RNA between benign (n= 7 [stones, stent changes for chronic pancreatitis, SOD]) and malignant bile samples (n=5 CC), there was no difference in relative expression CK19 (ΔΔCt 1.06 [95% CI 0.89-1.26]) or CD45 (ΔΔCt 1.06 [95% CI 0.66-1.69]).

RNA isolated from stents retrieved at ERCP had highly variable quantities of total RNA. More importantly, stent samples varied considerably with regards relative expression of CK19 and particularly CD45. Examples of data from two stent samples are shown in Figure 26 (other data not shown). This suggests that these samples are unlikely to be suitable for gene expression studies unless larger numbers of samples are used in order to reduce the variation seen between individual samples.

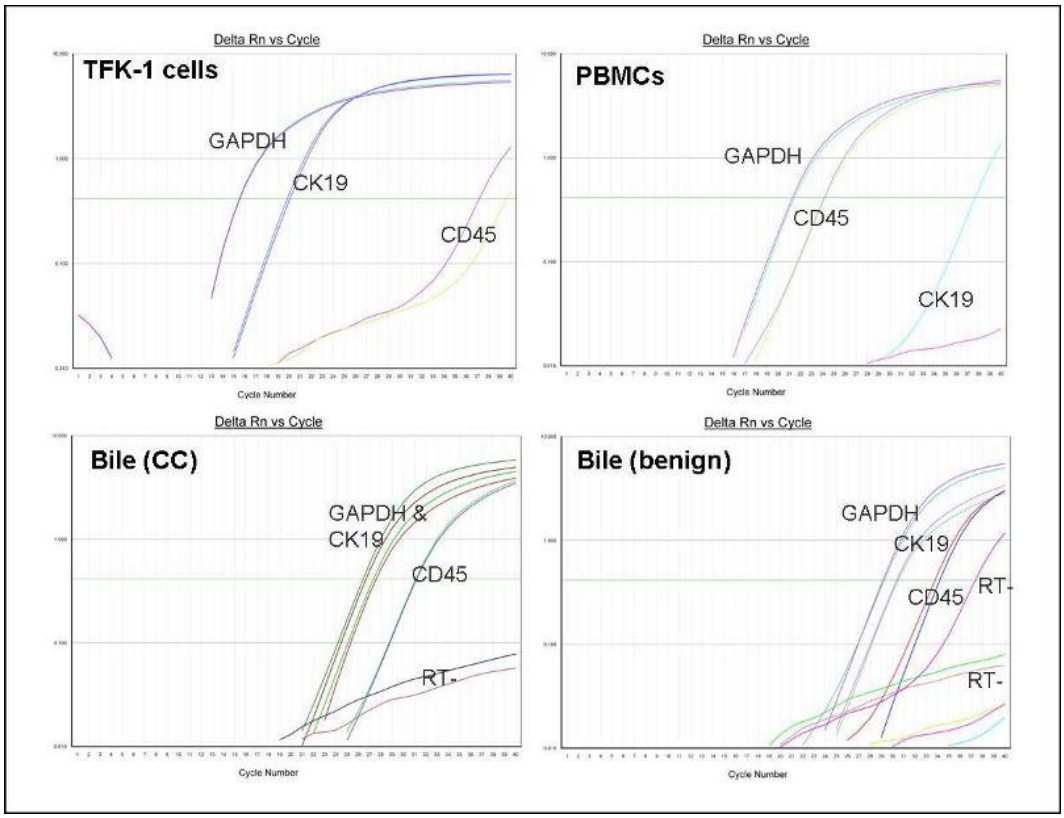


Figure 30. Representative qPCR plots for CK19 and CD45 as markers of relative expression of epithelial and leukocyte RNA respectively in bile samples. Note that in the BTC cell line (TFK-1) and PBMC controls, expression of the relevant markers are high (near to the GAPDH reference gene) or negative as exprected. In the clinical samples, the higher expression of CK19 suggests a greater contribution of epithelial to leukocyte RNA in bile, but with a lesser difference than that found in biliary brushings.

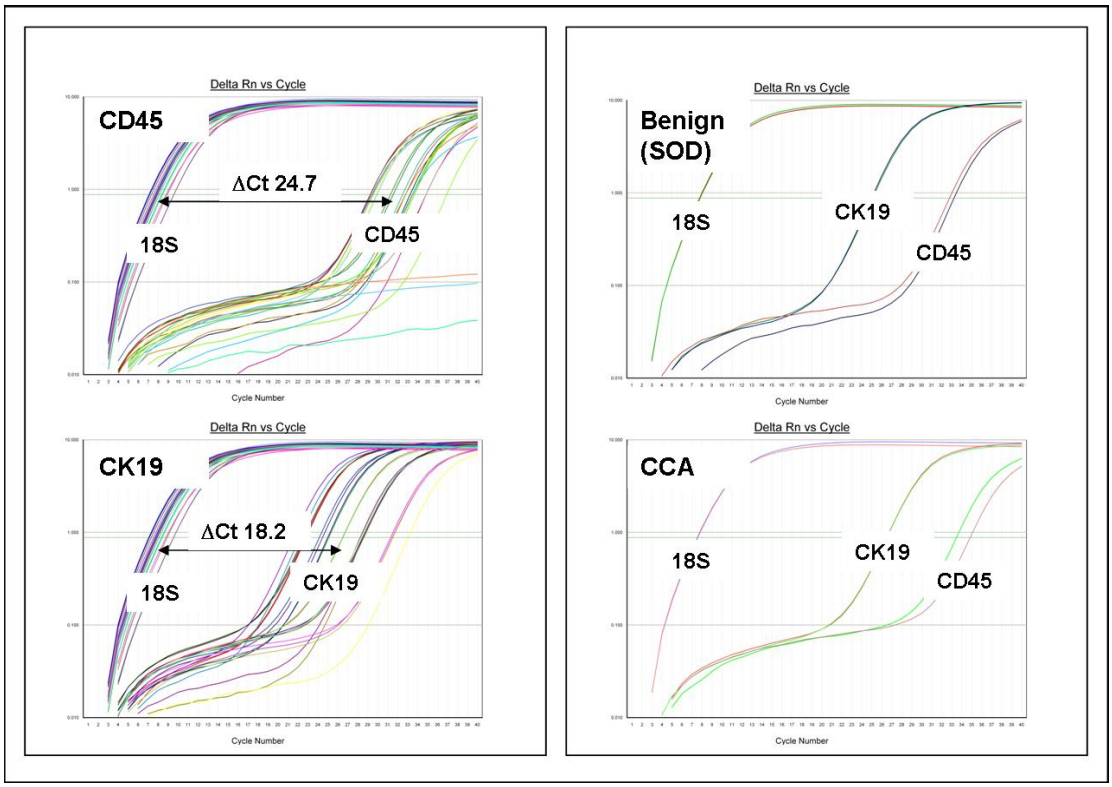


Figure 31. Left: Relative expression of epithelial (CK19) and leukocyte (CD45) RNA markers showing a greater proportion of epithelial cell RNA in bile. Right: Representative qPCR amplification plots from benign and malignant bile samples using the 18S reference gene showing that both have a much greater proportion of epithelial to leukocyte RNA.

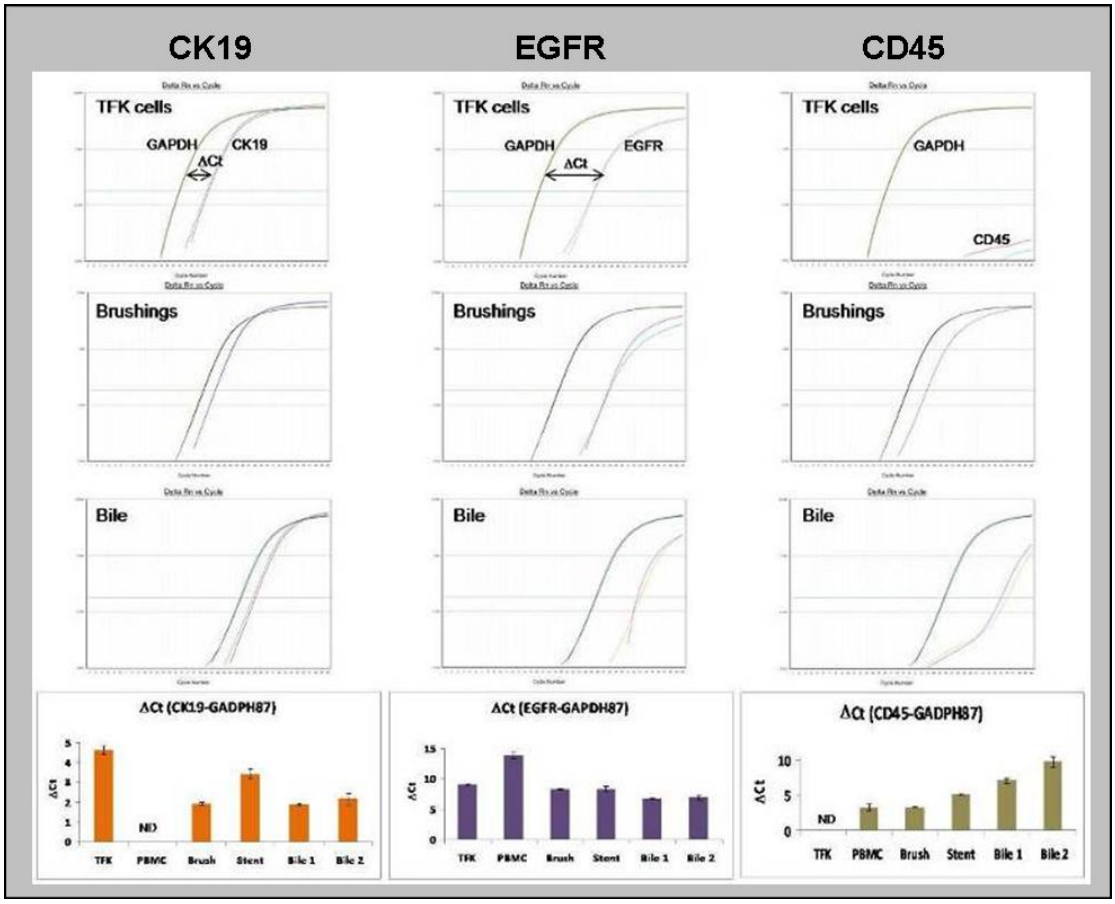


Figure 32. Use of the ΔCt method to compare gene expression in different sample types. qPCR plots are shown in the graphs and ΔCt values tabulated in bar charts below. Note that CK19 in the TFK-1 cell line has a slightly lower expression than the clinical samples despite being a pure cell line, suggesting an element of dedifferentiation and loss of full CK19 expression. Also, CD45 expression varies more than other genes tested.

4.4 Discussion

RNA isolated from clinical samples of bile and biliary brushings is shown to be highly degraded. However, application of methodology developed for similarly degraded RNA isolated from FFPE tissues allows gene expression profiling that has, as yet, rarely been applied to endoscopically obtained biliary samples. The clinical importance is primarily related to the fact that biliary samples obtained at ERCP are far more readily available than relatively rare surgical resection specimens. In addition, profiling of biliary brush cytology is likely to identify novel disease biomarkers in bile and/or blood that will assist clinical investigation and treatment in the future.

The *in vitro* experiments demonstrate that the methodology for RNA isolation is suitable and can provide high quality purified total RNA. In contrast to published data in biliary epithelial cell lines (Benedetti 1997, Ju 2002), our data suggest that the RNA degradation does not occur as a result of short term direct cell exposure to potentially noxious bile or x-ray contrast agents. These conditions appear to have little effect on the quality of RNA isolated from biliary epithelial cells cultured for up to 24 hours. We did not quantify the number of cells surviving in such environments and it is still possible that significant cell death occurred *in vitro*. However, if this were the case we would have expected to find significant amounts of degraded RNA, which was not found. These data also suggest that the RNA degradation occurs *in vivo* and as such is not amenable to methodological variations to prevent this. Maximisation of data acquisition is therefore dependent on the use of

appropriate cDNA synthesis, amplification and/or qPCR methods after isolation of the RNA.

A potential source of significant 'noise' and false positive results in biliary samples is the presence of leukocytes which are commonly found in patients with biliary disease and biliary obstruction. Relative gene expression of epithelial cell markers (CK19) and leukocyte markers (CD45) using the ΔCt method suggest that the primary source of RNA in our samples was epithelial in origin with a relative quantity many orders higher than that of leukocyte RNA. We are therefore confident that we have representative RNA expression data from our samples.

Our findings suggest that RNA isolated from bile and biliary brushings is suitable for gene expression profiling. This provides a potential route for development of new biomarkers for the diagnosis of malignant biliary strictures, as well as further investigate the biology of biliary diseases such as primary sclerosing cholangitis. Our aims are therefore to further explore gene expression in benign and malignant biliary disease using these methods along with whole genome RNA expression profiling using the microarray platform.

Chapter 5.

5 Whole genome RNA expression profiling of RNA isolated from biliary brushings

5.1 Introduction

We have demonstrated that RNA isolated from bile and biliary brushings is degraded and of relatively low quantity but can be used for gene expression analysis using qPCR. Similar quantities and quality of RNA isolated from other sources such as FFPE tissues, bronchial brushings and urine have been shown to be useful for whole genome expression profiling using microarray analysis (Lehmann and Kreipe 2001; Cronin, Pho et al. 2004; Mengual, Burset et al. 2006; Coudry, Meireles et al. 2007; Penland, Keku et al. 2007; Spira, Beane et al. 2007; Linton, Hey et al. 2008). One of these studies showed that differential gene expression using microarray analysis of bronchial brushings from macroscopically normal epithelium in the upper bronchus, identified tumour at a distant site in the bronchial tree (Spira, Beane et al. 2007). This suggests that a 'field change' of alterations in gene expression exists and this phenomenon has also been proposed in the unaffected liver of patients with PSC and distant cholangiocarcinoma (CC) (Fleming, Boberg et al. 2001). A major problem with biliary brush cytology for

the diagnosis of CC is the low sensitivity even when directly brushing malignant strictures. The work by Spira et al in bronchial brushings and the low sensitivity of biliary cytology provides an obvious rationale to evaluate this approach in biliary brushings.

Microarrays are referred to by a number of different names by manufacturers and in the literature (microarrays, DNA microarrays, RNA expression arrays, gene expression arrays, oligonucleotide spotted arrays, gene chips etc). All refer to products with similar techniques and aims. The principle of microarrays is that a particular mRNA sequence will bind to a complementary DNA template from which it was made, and binding to a cDNA oligonucleotide can be quantitated, often using immuno-fluorescence. Microarrays have been developed using standardised information on gene sequences and nomenclatures from the Human Genome Project and public databases such the National Center for Biotechnology Information (NCBI) databases. This has allowed mRNA specific DNA oligonucleotides to be designed, synthesized and spotted densely onto glass plates in the gene chips. The confusion in nomenclature between RNA expression and DNA microarrays arises from the fact that cDNA oligonucleotides complementary to mRNA are used to bind to the oligonucleotide probes on the chips and hence identify gene expression.

Most commercially available microarray chips now include genes spanning the entire genome as described by NCBI databases. The signal (e.g. colour or immuno-fluorescence) released by bound cDNA can be measured as relative abundance for each data point on the chip using a high resolution

laser, and the vast amounts of data generated are analysed using software designed for specific questions such as quantification of gene expression, identification of DNA mutations or splice variants. Worldwide agreements on publications using microarray data, set out as the microarray mark-up language (MAML), require that all raw source files and experimental methods are submitted for public use in one of the main microarray depositories such as the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). In time, this will provide an enormous resource for data analysis that will be available for work such as biomarker development and improved understanding of cell biology in health and disease. Microarray technology has been successfully applied to the identification of biomarkers in other cancers including pancreas, breast, prostate and lung cancers (Perou, Sorlie et al. 2000; Dhanasekaran, Barrette et al. 2001; Garber, Troyanskaya et al. 2001; Iacobuzio-Donahue, Maitra et al. 2002).

Searches of databases reveal only very limited microarray data sets in BTC to date, in line with the limited research in this field. Studies have been performed using BTC cell lines and surgical resection specimens from partial hepatectomy for cholangiocarcinoma (Obama, Ura et al. 2005). Cancer cell lines provide useful data with regards to cancer biology and alteration in cellular pathways in response to treatment interventions but do not necessarily translate into clinical practice. Microarray analysis using surgical resection specimens has been mainly performed using samples of intrahepatic cholangiocarcinoma (Obama, Ura et al. 2005; Jinawath, Chamgramol et al. 2006; Hass, Nehls et al. 2008). The same group published

two of these papers investigating the expression profile of intra-hepatic CC with or without evidence of liver fluke infection. They used laser capture microdissection (LCM) to separate out biliary epithelial tissue and identified a list of differentially expressed genes in intrahepatic CC. Although the addition of LCM should improve the accuracy of gene expression profiling, these cases do not necessarily represent the different pattern of extrahepatic CC more commonly seen in the Western world.

A group in Germany published data on differential gene expression between blocks (n=10) of fresh frozen intrahepatic CC compared to the surrounding "normal'' liver (Hass 2008). This study used the first generation of Affymetrix U133A GeneChips and identified 552 differentially expressed genes in CC. The most significant up-regulated gene in this study was osteopontin with a fold change in cancer versus benign disease of 33 ($p=<0.001$). However, the study was limited by the fact that adjacent liver tissue will greatly differ in cellular content (predominantly mature hepatocytes) from that of CC (predominantly adenocarcinoma and stromal cells), and these tissues are therefore not directly comparable.

One paper has been published using fresh surgical samples of resected extrahepatic BTC (Hansel, Rahman et al. 2003). This group used scrapings of biliary epithelium resected at the time of Whipple's procedures, whole surgical resections of BTC (mostly gallbladder cancer) and BTC cell lines analysed using the first generation of Affymetrix chips (HG U133A). Although these samples are likely to be more representative than the intrahepatic tumour

samples investigated by others, data are still confounded by the use of cell lines, stromal tissue and predominance of gallbladder material, and hence are not ideal for analysis. However, a list of 282 genes with greater than 3 fold change in expression (up- or down-regulated) was generated, many of which have previously been shown to be altered in BTC.

5.2 Aims

The main aim of the work presented in this chapter was to use whole genome RNA expression profiling of RNA isolated from biliary brushings of benign and malignant biliary disease for the identification of genes that may be potential biomarkers for the diagnosis and prognosis of BTC.

Secondary aims were:

- To identify genes that may be important in the biology of BTC.
- To assess whether microarray of biliary brushings is feasible for investigating gene expression in other biliary diseases hampered by paucity of clinical material such as PSC.
- To assess whether there may be a role for expanding the work on biliary brushings in order to identify multiple gene signatures in the form of ‘diagnostic chips’ that may be useful in diagnosis and prognosis of malignant biliary strictures.

5.3 Materials and methods

5.3.1 Biliary brush samples

Brushings of biliary epithelium were obtained from macroscopically normal bile duct wall, and benign and malignant strictures using a wire guided, sheathed endobiliary brush (Boston Scientific, Notick, USA) at the time of clinically indicated ERCP. Ethical approval was granted by the joint UCL/UCH ethics committee for sample collection and research (ref 06/Q0152/106) with separate site specific assessment at the Royal Free Hospital. Details of methods for RNA isolation are described in Chapter 4.

Clinical samples used for the microarray analysis are described in Table 12. Definitive diagnosis was reached in all by cytology or histology confirmatory of cholangiocarcinoma. Those with sphincter of Oddi dysfunction (SOD) were considered to have normal biliary epithelium and those with benign findings had benign clinical history, benign cytology and no evidence of malignant disease after at least 12 months follow up. In 5 out of the 10 cases, diagnostic cytology was sent for pathological assessment at the same time a second sample was taken for research. In the remaining 5 cases, a diagnosis was already clear and no clinical cytological samples were obtained at the time research samples were taken.

Table 12 Summary of biliary brush samples used for microarray analysis

Sample	Diagnosis	Total RNA	Agilent RIN	Malignant cytology	Biliary sepsis	Further comments
B1	Benign stricture (chronic pancreatitis)	1123	5.1	No	No	Stent in situ at time of ERCP
B2	Biliary SOD	150ng	4.3	Not done	No	Normal ERCP with elevated biliary sphincter pressure
B3	Pancreatic SOD	1657ng	2.5	Not done	No	Normal ERCP with elevated pancreatic sphincter pressure
B4	Benign stricture (chronic pancreatitis)	2840ng	1.8	No	No	Stent in situ at time of ERCP.
C5	Low CBD CC	270ng	5.8	Not done	No	Stent change
C6	Hilar CC	2290ng	2.9	Not done	No	Elevated CRP (138) but no other evidence cholangitis
C7	Hilar CC	690ng	2.5	Not done	No	Stent change
C8	CBD CC	2400ng	2.0	Adenocarcinoma	No	Diagnostic ERCP
C9	Hilar CC	650ng	1.8	Suggestive but not diagnostic	Yes	Pus in biliary tree but normal clinical parameters
C10	Hilar CC	815ng	2.5	Adenocarcinoma	No	Diagnostic ERCP

The biliary brushings were immediately snap frozen in liquid nitrogen within a few minutes of collection. The tip of the brush was cut off into a 2ml Eppendorf tube using a wire cutter. The tube was then immersed and transported in liquid nitrogen and transferred to a -80^{0}C freezer until further processing.

5.3.2 RNA isolation and purification

Total RNA was isolated and purified as described in chapter 4. In brief, RNA was extracted from the biliary brush samples using phase separation and ethanol precipitation using TRI Reagent (Ambion) as per the manufacturer's protocol. The total RNA solution was then purified using DNase digestion (4U TurboDNase, Ambion) and spin column filtration steps (RNEasy MinElute,

Qiagen). Total RNA was stored in 12μl volumes of RNase free water at -80^{0}C until further use for the microarray experiments.

5.3.3 Whole genome RNA expression profiling using Affymetrix U133 plus 2.0 GeneChips

The RNA amplification, labelling and hybridisation to the Affymetrix GeneChips were performed as a paid microarray service by UCL Scientific Support (Catherine King, Wolfson Institute for Biomedical Research, UCL).

5.3.3.1 Choice of microarray chips

Many commercially available microarray chips are now marketed. The choice of chip depends on the information one aims to obtain and the local facilities available. Each chip has strengths and weaknesses and the latest generation are usually designed for specific purposes. For example, the Affymetrix Human Exon Arrays have probes designed to detect each exon of a gene and can therefore detect splice variants for specific genes. However, they require larger starting quantities of RNA and are less efficient in detecting small mRNA sequences with few exons.

The microarray chips used were chosen following a review of the literature on RNA from FFPE samples and advice from the microarray service at UCL. Different manufacturers use different size oligomer probes. Affymetrix uses smaller 25mer probes and the Agilent systems use 60mer probes with no

difference in accuracy for gene expression. In view of the degraded RNA and the demonstration in earlier work (Chapter 4) that PCR amplicon size is important for quantification of gene expression, it was felt that the Affymetrix arrays were more likely to provide accurate results. When assessing overall gene expression, a study comparing the main Affymetrix chips concluded that, despite variation in probe level expression, there was little variation in gene level expression of differentially expressed genes between chips (Robinson 2007).

The chip chosen was the Affymetrix Human U133 plus 2.0 GeneChip (Figure 33) which has been the most widely used chip to date. This chip has a smaller probe level variance than other Affymetrix chips. Unlike the exon array, it cannot detect splice variants for each gene and detects overall expression of each gene using multiple 25mer probes per gene. The level of expression is determined by the number of bound probes per gene.

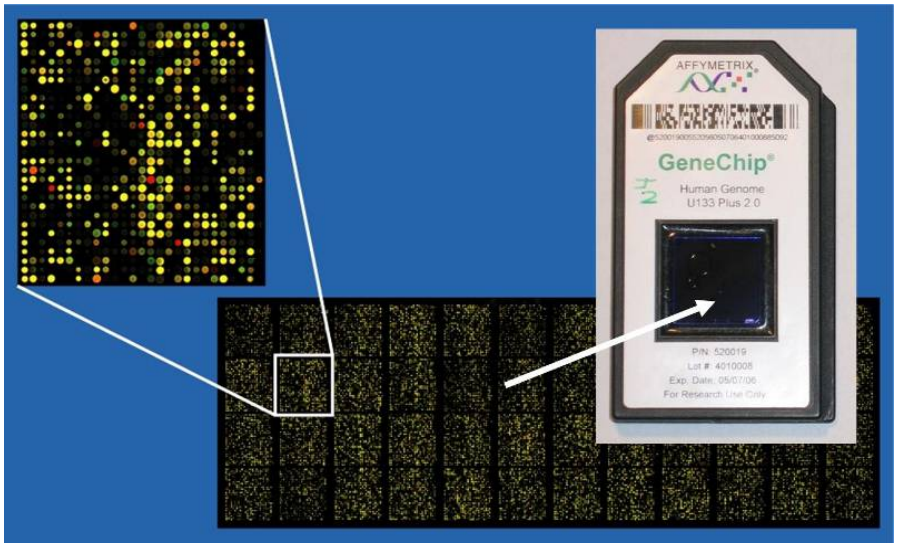


Figure 33 Affymetrix GeneChip showing magnified areas of the oligonucleotide spotted plate and bound immunofluorescence tagged sample cDNA measured by the chip reader. (Collage of images taken from Affymetrix.com

5.3.3.2 Statistical aspects of sample numbers for microarray experiments

Advice for experimental design was given by staff at the microarray service (Dr Mike Hubank, Institute of Child Health) and the Bloomsbury Centre for Bioinformatics (Dr Sonia Shah). In general, the larger the number of samples, the more reliable the results and the more likely it is that small differences will be identified. Our work centres on biomarker development rather than cell biology and hence we were interested in identifying only larger differences between groups. The advice was that the minimum sample size required for application of statistical methods to gain meaningful results is 3 samples per group. However, microarray experiments are expensive and clinical samples are precious. Therefore we first undertook a pilot experiment using only 2 samples per group to assess feasibility of the work. These experiments suggested good quality data and identified genes already implicated in BTC so further supporting the technique. The variation in level of gene expression was smaller in the benign group. We therefore concluded that the addition of six further samples would likely be sufficient in order to detect genes with large differential expression for biomarker development. In total, 6 samples were taken from patients with malignant disease and 4 from patients with benign biliary strictures.

5.3.3.3 RNA amplification and labelling

In a deviation from standard protocol, we did not use the Affymetrix RNA labelling kits prior to hybridisation to the Affymetrix chips and used alternative kits more appropriate for the degraded nature of our RNA. Both the steps described have been tested and approved by the manufacturer and others for use on the Affymetrix microarray platform.

The WT-Ovation FFPE RNA Amplification system (Nugen, USA) uses a mixture of random and oligo dT primers to synthesize amplified cDNA. Like many systems, transcription is initiated at the 3' end of the RNA using poly dT primers but also uses the addition of random priming of the whole transcriptome. This significantly improves amplification and cDNA synthesis using short segments of degraded RNA, where degradation preferentially occurs at the polyA end of RNA, and is a more efficient RNA amplification and cDNA synthesis method for degraded RNA isolated from FFPE for microarray studies. The cDNA synthesis step results in amplified levels of cDNA (up to 10μg) complementary to mRNA sequences. The starting quantity of total purified RNA used was 75ng. Link to the instruction manual:
http://www.nugeninc.com/tasks/sites/nugen/assets/File/user_guides/userguide_wt_ov_ffpe.pdf

cDNA labelling for hybridisation to the microarray chips was performed using the FL-Ovation cDNA Biotin Module v2 (Nugen, USA) as per the manufacturers instructions:

(http://www.nugeninc.com/tasks/sites/nugen/assets/File/user_guides/userguide_fl_ov_biotin.pdf).

The process uses a thermal cycling fragmentation step to produce single stranded fragments (50-100bp) of cDNA followed by enzymatic attachment of a biotin labelled nucleotide to the cDNA. The biotin label is used for quantification by the Affymetrix chip readers. Prior to this step, samples of amplified cDNA were assessed using the Nanodrop ND-1000 spectrophotometer to quantify 5μg of required input cDNA and to ensure 260/280 ratios were over 1.8.

5.3.3.4 Hybridisation to DNA microarray chips

Processing of GeneChips was performed as a paid service by the microarray facility, Scientific Support, Wolfson Institute for Biomedical Research, UCL. Each sample of biotin labelled cDNA was hybridised to single Affymetrix human U133 plus 2.0 GeneChips and processed as per the manufacturer's instructions. GeneChip signal intensities were recorded using a GeneChip Scanner 3000 (Affymetrix) and raw probe level data stored as standard Affymetrix CEL (.cel) data files for data analysis.

5.3.4 Microarray analysis

Statistical analysis and assessment of quality control in our samples was performed as a paid service by Dr Sonia Shah at the Bloomsbury Centre for Bioinformatics.

Prior to analysis of gene expression data, the raw data obtained from the CEL data files requires a number of quality control and normalisation steps to ensure that quantities and quality of data from each chip is similar and comparable. Variability in biotin signal intensity can be normalised using normalisation genes and in built chip signal controls so that each chip provides comparable data. However, variation in quality of the data recovered cannot be easily normalised and therefore, samples with significant quality control variance from other samples must be removed from the pool before data analysis. In general, consistency in quality of input RNA is more important than the actual quality of RNA at baseline. Therefore, if we had samples of good quality starting RNA, they would not be suitable for inclusion with our mostly degraded samples. As shown above, none of the samples were of good quality RNA and we chose samples with the least degraded electropherogram plots.

Each gene has multiple probes with separate signals. Summarisation combines the individual probe intensities to calculate single gene intensity used to calculate gene expression. Background correction adjusts the value of each probe signal to account for background hybridisation. The data were normalised using the Affymetrix microarray suite (MAS5) algorithm. During MAS5 background correction, the chip is divided into regions (default=16). The lowest 2% of signals are used to compute the background for that region. Each probe intensity is adjusted based upon a weighted average of each of the background values. Perfect match (PM) probes are adjusted using mismatched (MM) signals and the probe signals are then summarised into

gene signals using the Tukey Biweight algorithm. The data were lastly normalised by scaling all arrays to have the same mean intensity (Figure 34).

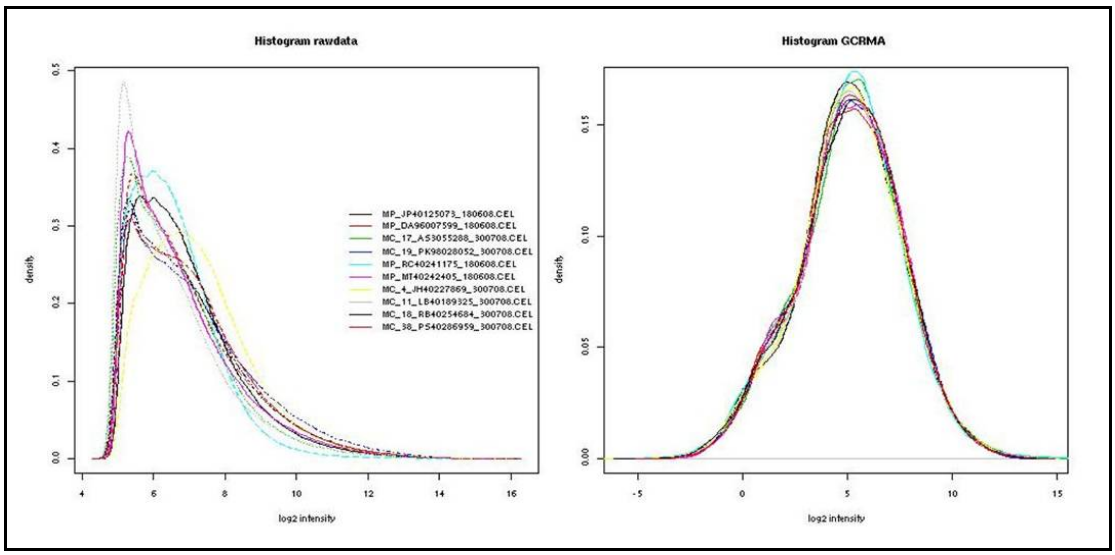


Figure 34 Histogram of individual GeneChip signal intensities before and after MAS5 normalisation showing very similar profile for all samples

5.3.5 Quality control

Quality control of the raw and normalised data was performed using the following methods:

5.3.5.1 Box plots

The box plots (Figure 35) display the median, and upper and lower quartiles of overall intensities for each sample. The whiskers show data 1.5x the inter-quartile range. Figures outside this range (shown as dots) are considered outliers and provide data on those samples with a higher variability in signal intensity. One sample (RC) had significantly more outliers than the others and was considered unusual, marking it for possible removal from data analysis.

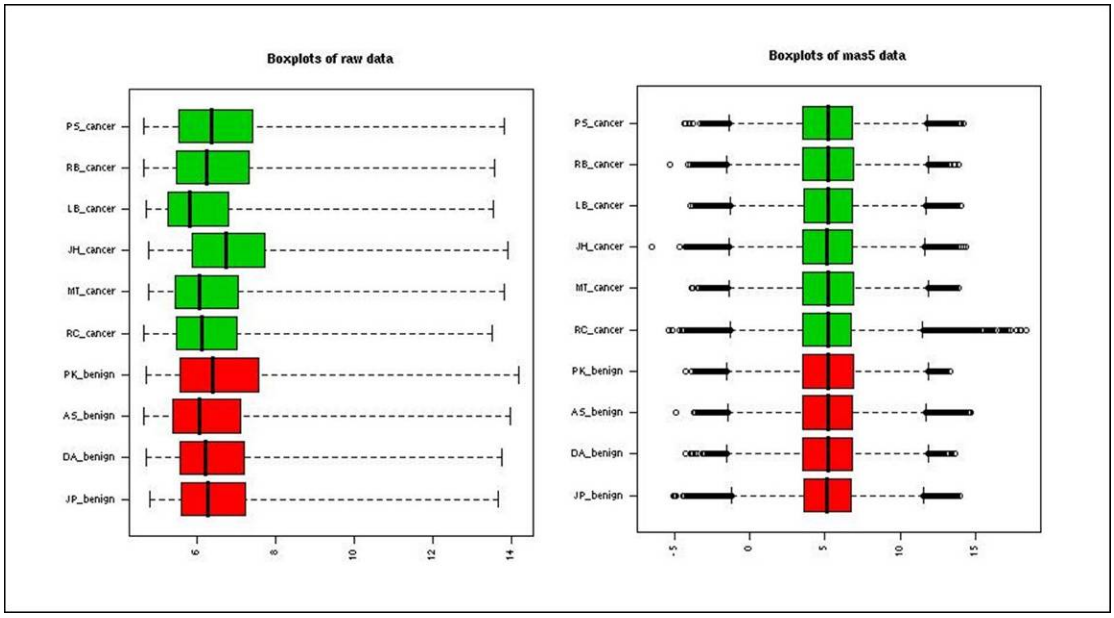


Figure 35 Box plot of microarray signal intensities before and after normalisation

5.3.5.2 Histograms of PM versus MM probe signal intensities

Affymetrix GeneChips consist of a combination of mismatch (MM) and perfect match (PM) probe sets. The MM probes measure the degree of non-specific hybridisation while PM measure specific hybridisation. The cRNA should bind the MM probes less strongly than to the PM probes and have a weaker intensity signal than the PM probes. If the PM curve is not significantly different from the MM curve, this implies a low signal to noise ratio. In such cases it is often difficult to distinguish truly differentially expressed genes from the noise. In our samples, the PM versus MM plots intensities differed as expected but had some variation in distribution between plots (Figure 36).

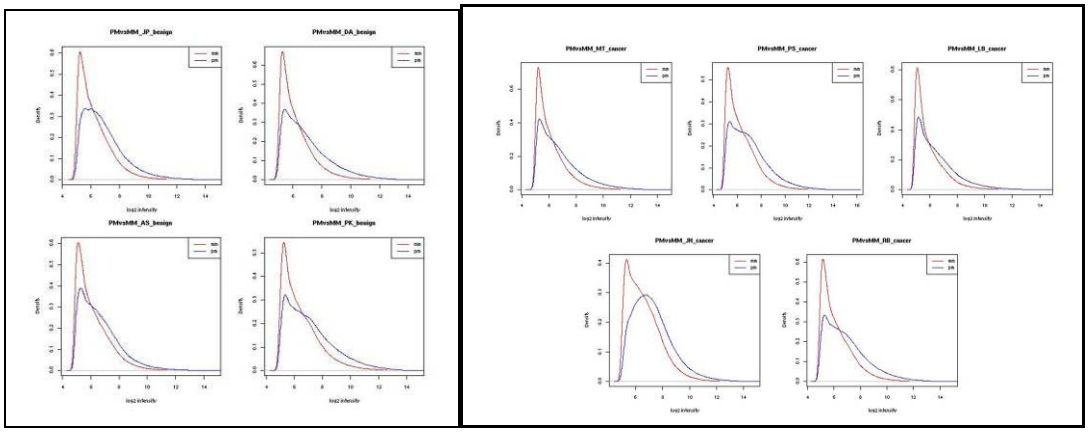


Figure 36 Histograms of PM (perfect match) versus MM (mismatch) probe signal intensities showing similar strength and spread of signal intensities for benign (left) and malignant (right) disease samples

5.3.5.3 RNA degradation plots

RNA degradation occurs preferentially at the 5' end of the molecule. Genes and transcripts are represented on Affymetrix chips as a series of oligonucleotide probes numbered 0-10, with probe 0 being the most 5' sequence and probe 10 the most 3' sequence. RNA degradation plots show expression as a function of 5'-3' position of probes. The plot in Figure 37 shows the average intensity of the probes classified by this order. Each line corresponds to an array and the slope of its trend indicates potential RNA degradation and/or inefficient labelling.

The most important factor when assessing quality is consistency, and so it may be acceptable for all samples to be of poor quality as long as they were similarly poor. One sample (RC) had an RNA degradation plot that was much flatter than the others. This suggested either much better quality RNA (unlikely) or very poor RNA with little hybridisation to any of the probes.

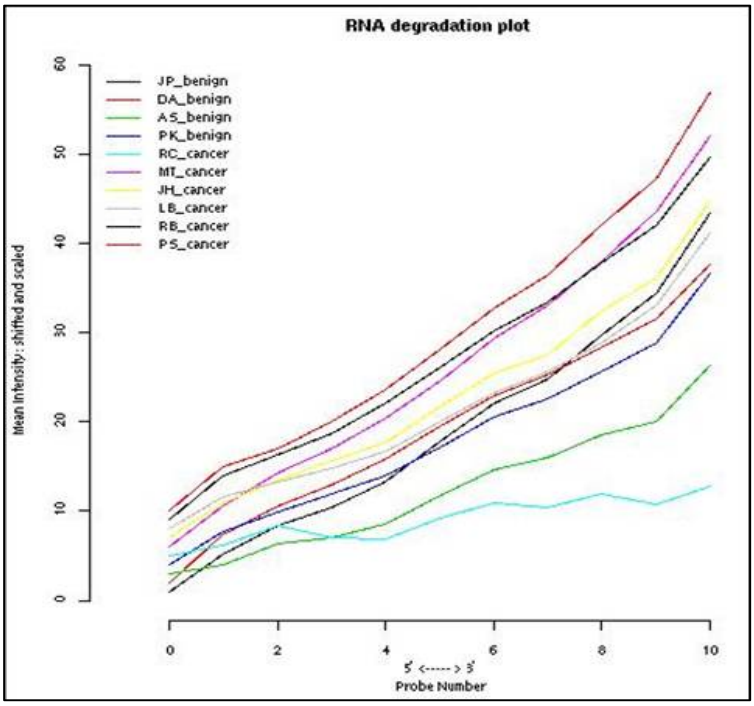


Figure 37 RNA degradation plots for each GeneChip sample showing slope of curve as a feature of hybridisation to probes at different points along the length of control genes.

5.3.5.4 SimpleAffy software

SimpleAffy was used as a second commonly used quality control step. The main data analysis tests are as follows:

Number of genes called 'present'

All samples were recorded as having over 50% call rates, considered to be satisfactory even when using good quality RNA.

3' to 5' ratios

Unlike the multiple probe sets analysed in the RNA degradation plots above, SimpleAffy assesses only β-actin and GAPDH, two relatively long genes. By

comparing the intensity values from the 3' probe set to the 5' or mid-point probe sets a 3':5' ratio can be calculated. A high ratio indicates significant RNA degradation or a poor *in vitro* transcription step. GAPDH 3':5' ratios are plotted as circles; ratios > 1 are coloured red (unacceptable). B-actin 3':5' ratios are plotted as triangles. Being longer than GAPDH, unacceptable ratios are > 3 and also coloured red. Acceptable ratios for both are coloured blue. Many of our samples had high ratios, most notably sample RC (Figure 38).

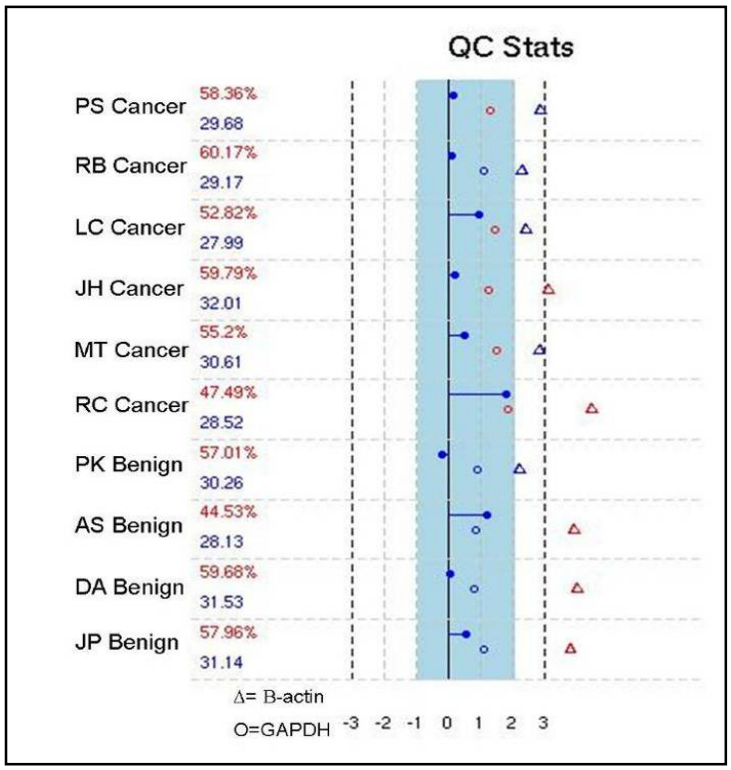


Figure 38 SimpleAffy quality control using GAPDH and beta-actin normalisation genes showing variability in signal intensity

5.3.5.5 Correlation plots

The correlation plot is a map of the array-array Spearman rank correlation coefficients. Self-self correlations run diagonally and by definition have a correlation coefficient of 1.0. Data from similar tissues or treatments should

have higher coefficients. The plot can be used to detect outliers. Sample RC had very different gene expression values from the rest of the samples which is reflected in the lower correlation to other samples (Figure 39).

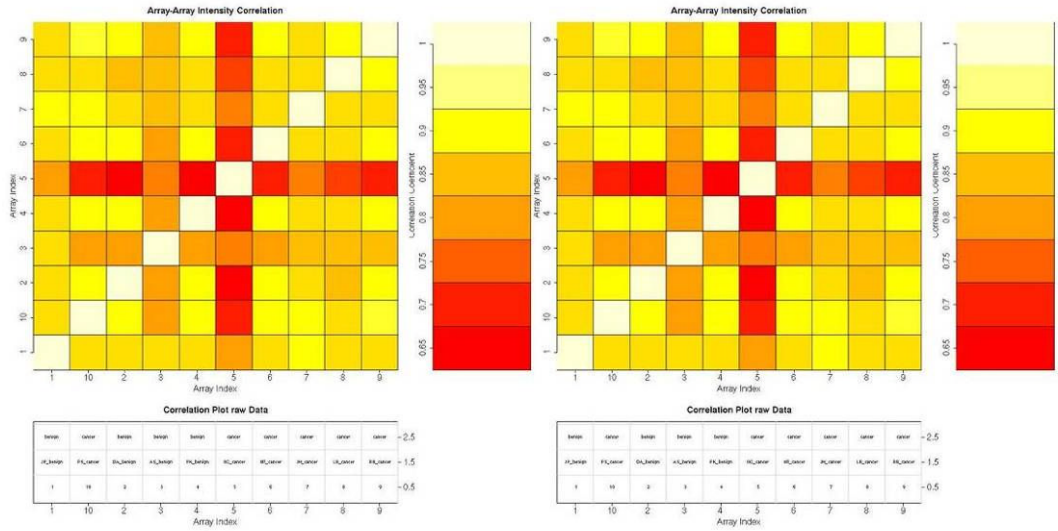


Figure 39 Spearman correlation plots for each GeneChip showing marked variation in one (5), moderate variation in one (3) and good correlation in the remaining samples.

5.3.5.6 Microarray data quality control; conclusions

Analysis of the raw and normalised data suggested a satisfactory level of quality for all samples except one (RC), which had many parameters outside those for the other samples. This sample was removed prior to further data analysis. The remaining samples had evidence of RNA degradation but were considered to have similar overall quality individually and between the benign and malignant groups and were therefore suitable for comparison and further data analysis.

5.3.6 Statistical considerations

5.3.6.1 LIMMA analysis for calculation of fold change in gene expression

Analysis of microarray data is performed using specially designed data analysis software. Our data were analysed by Dr Sonia Shah, Bloomsbury Centre for Bioinformatics, using the linear models (LIMMA) Bioconductor software designed for gene expression analysis. The software is based on a modified t-test and stabilises data for small sample numbers and allows for the presence of biological and experimental variation. Bayesian statistics are applied as a means of measuring the probability of an outcome (i.e. significant fold change) with the prior assumption of a null hypothesis.

5.3.6.2 Correction for multiple comparison testing

Statistical significance for most experimental work is usually set at a level of 5%. When using very large numbers of comparisons, such as the thousands used in microarray experiments, a significant number of false positive and false negative results will inevitably occur. Statistical techniques have been developed in order to reduce both type I and type II errors in analysis of microarray experiments. The method used for correction for multiple hypothesis testing in our work was the Benjamini-Hochberg (FDR) test using a p value cut-off of 0.05. This means that only 5% of the resulting list of differentially expressed genes are likely to be false positive results (i.e. 56

genes) as opposed to 5% of the entire probe set of 47,000 genes i.e. 2350 genes.

5.3.7 Gene expression data parameters used for gene listing

Genes that were absent in three or more samples per group were considered absent and excluded from analysis. For cell biology data, a fold change of 1.5 is usually set as the cut-off for alteration in gene expression. In order to attempt to identify biomarkers with a high capacity to differentiate benign from malignant disease groups, we used a fold change cut-off of 2 or more (measured as a log2 fold change in data files and gene expression plots). A statistical cut-off was set at a p value of 0.05. The resulting list of up- and down-regulated genes therefore only includes genes with a fold change of 2 or more with a p value of less than 0.05 after correction for multiple hypothesis testing.

5.4 Results: microarray gene expression data

In view of the size of the files and for ease of reference, results were recorded in the form of an HTML document which was also converted to Microsoft Excel format for searching of genes. The files and information contained within are too large to be included but the number of up- and down-regulated genes identified are summarised in Table 13.

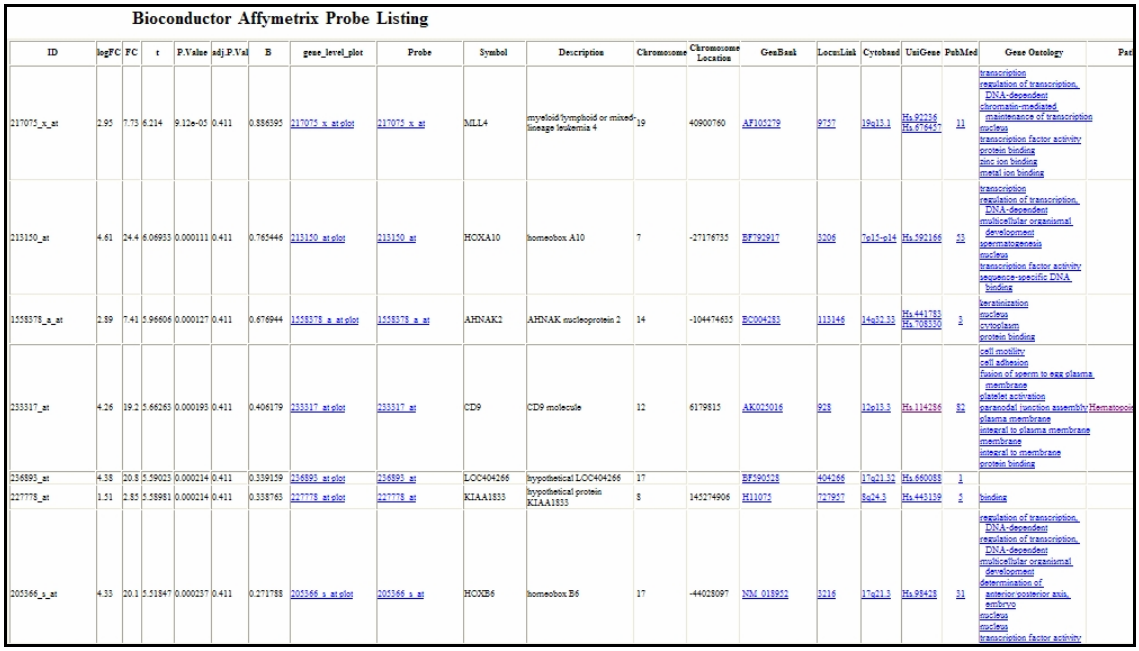
Table 13 Summary of numbers of genes with significant up- or down-regulation in microarray data

Expression pattern	Number of genes identified
Down-regulated	1001
No significant change	34057
Up-regulated	1141

5.4.1.1 HTML documents of altered gene expression

The HTML format provides an easy format to access the large amount of data in the results. In addition, the table of data includes other important and useful information such as the Affymetrix probe identification, fold change, p value, official gene symbol and name, gene functions and locations, chromosome location and links to NCBI databases including GeneBank, LocusLink and PubMed links for each gene. In addition, each gene listed has a link to the graphical display of the gene expression levels for each sample in each group so that the pattern and spread of gene expression can be viewed easily. These graphs are known as gene plots and examples are shown in Figure 41.

An image of the HTML format of a small section of the up-regulated genes is shown in Figure 40.



Bioconductor Affymetrix Probe Listing

ID	logFC	FC	t	P.Value	adj.P.Val	B	gene_level_plot	Probe	Symbol	Description	Chromosome	Chromosome Location	GenBank	LocusLink	Cytoband	UniGene	PubMed	Gene Ontology	Pat
217075_x_at	2.95	7.73	6.214	9.12e-05	0.411	0.886395	217075_x_at plot	217075_x_at	MLL4	myeloid/lymphoid or mixed-lineage leukemia 4	19	40900760	AF105279	9757	19q13.1	Hs.89336 Hs.87452	11	transcription regulation of transcription, DNA-dependent chromatin-mediated maintenance of transcription nucleus transcription factor activity protein binding zinc ion binding metal ion binding	
213150_at	4.61	24.4	6.06933	0.000111	0.411	0.765446	213150_at plot	213150_at	HOXA10	homeobox A10	7	-27176735	BF792917	3206	7p15-p14	Hs.592166	53	transcription regulation of transcription, DNA-dependent multicellular organismal development spermatogenesis nucleus transcription factor activity sequence-specific DNA binding	
1558378_a_at	2.89	7.41	5.96606	0.000127	0.411	0.676944	1558378_a_at plot	1558378_a_at	AHNAK2	AHNAK nucleoprotein 2	14	-104474635	BC004283	113146	14q32.33	Hs.441783 Hs.708350	3	keratinization nucleus cytoplasm protein binding	
233317_at	4.26	19.2	5.66263	0.000193	0.411	0.406179	233317_at plot	233317_at	CD9	CD9 molecule	12	6179815	AK025016	928	12p13.3	Hs.114286	82	cell motility cell adhesion fusion of sperm to egg plasma membrane platelet activation paranodal junction assembly plasma membrane integral to plasma membrane membrane integral to membrane protein binding	Hematopoi
236893_at	4.38	20.8	5.59023	0.000214	0.411	0.339159	236893_at plot	236893_at	LOC404266	hypothetical LOC404266	17		BF590528	404266	17q21.32	Hs.660088	1		
227778_at	1.51	2.85	5.58981	0.000214	0.411	0.338765	227778_at plot	227778_at	KIAA1833	hypothetical protein KIAA1833	8	145274906	H11073	727957	8q24.3	Hs.443139	4	binding	
205366_s_at	4.33	20.1	5.51847	0.000237	0.411	0.271788	205366_s_at plot	205366_s_at	HOXB6	homeobox B6	17	-44028097	NM_018952	3216	17q21.3	Hs.98428	31	regulation of transcription, DNA-dependent regulation of transcription, DNA-dependent multicellular organismal development determination of anterior/posterior axis, embryo nucleus nucleus transcription factor activity	

Figure 40 HTML format of microarray data presentation showing a small section of the table.

5.4.2 Choice of candidate genes for further validation using qPCR

Initial examination of the microarray data of up- and down-regulated genes appeared to identify many genes with different functions, locations and from different pathways and cell types. Selection of candidate markers requires exhaustive searches of databases to understand the biological function and plausibility of each gene to assist with decision making

Examination of the data appears to include a significant number of leukocyte genes which raises the question that there may be significantly more leukocyte contamination in the cancer samples compared to the benign samples as a result of cholangitis in patients with complex biliary strictures. However, further exploration of the function and expression of these genes shows that they have usually already been shown to be expressed by other tissues including epithelial cells (e.g. lymphoid enhance binding protein 1 [LEF1], transforming growth factor B1 [TGF-B1], Interleukin receptor kinase 3 [IRKAM]). An important finding providing evidence against a leukocyte bias in the cancer samples is the lack of leukocyte specific genes such as CD45, leukocyte specific protein-1 (LSP1), leukosialin (CD43), CD18 (MHM23), cathepsin G, leukocyte alkaline phosphatase (LAP), CD11 and CD166. These genes are commonly used as leukocyte specific markers for cell sorting and other work and do not appear in either the up- or down-regulated list of genes.

Selection of a shortlist of genes was made using a number of criteria discussed in turn:

5.4.2.1 Up- or down-regulated expression

Even though the down-regulation of genes is clearly important in cancer biology, the main aim of the project was identification of biomarkers. It is unlikely that down-regulated genes will directly code for easily measurable proteins in blood or bile and so are less likely to be useful for biomarker identification than up-regulated genes. Identification of altered protein expression up- or down-stream from down-regulated genes would require further complex data analysis of inter-related pathways with the possibility of introduction of further error. We therefore chose to concentrate on up-regulated genes for the shortlist.

5.4.2.2 Level of alteration in gene expression (fold change)

The higher the fold change between disease and controls, the more likely that the protein product will have significant differences between groups. However, it is also possible that genes may be up-regulated as a result of mutation in that gene which may be non-coding or code for biologically inactive proteins which may not be measurable using standard commercial assays such as ELISA. Another factor affecting protein product is the half life of the protein so that genes coding for proteins with a short half life may result in limited fold change in the measurable protein product. Alternatively, some genes with smaller fold change in expression may code for proteins with a long half life

and result in considerable differences in protein levels between groups. The cut-off for fold change used in our microarray data analysis was 2 which will exclude many biologically important genes but was chosen to limit the number of genes to manageable levels.

5.4.2.3 Level of statistical significance

Despite correction for multiple hypothesis testing being performed as part of the microarray data analysis, some of the gene expression differences will not be real alterations. Using a p value of 0.05 will still result in 5% of the genes being falsely identified as different between groups. The stronger the statistical significance in fold change, the more likely the difference is not related to chance.

5.4.2.4 Overlap of gene expression between benign and malignant groups

The gene expression plots generated by the microarray analysis software (gene plots) provide a visual display of gene expression for each sample in each group as well as displaying the mean level for each group. Some genes have no overlap in expression between cancers and benign samples raising the possibility that protein levels will similarly be altered in such a way that a definite positive and negative result may be possible. Other genes have significant differences between groups but each group may have a wide variation in expression level and overlap with the other group. This is more likely to result in wide spread and overlap in protein expression such as that

seen with CA19-9. In addition to the overlap, small variations in expression within a group are likely to be more reliable than those with a large variation in expression.

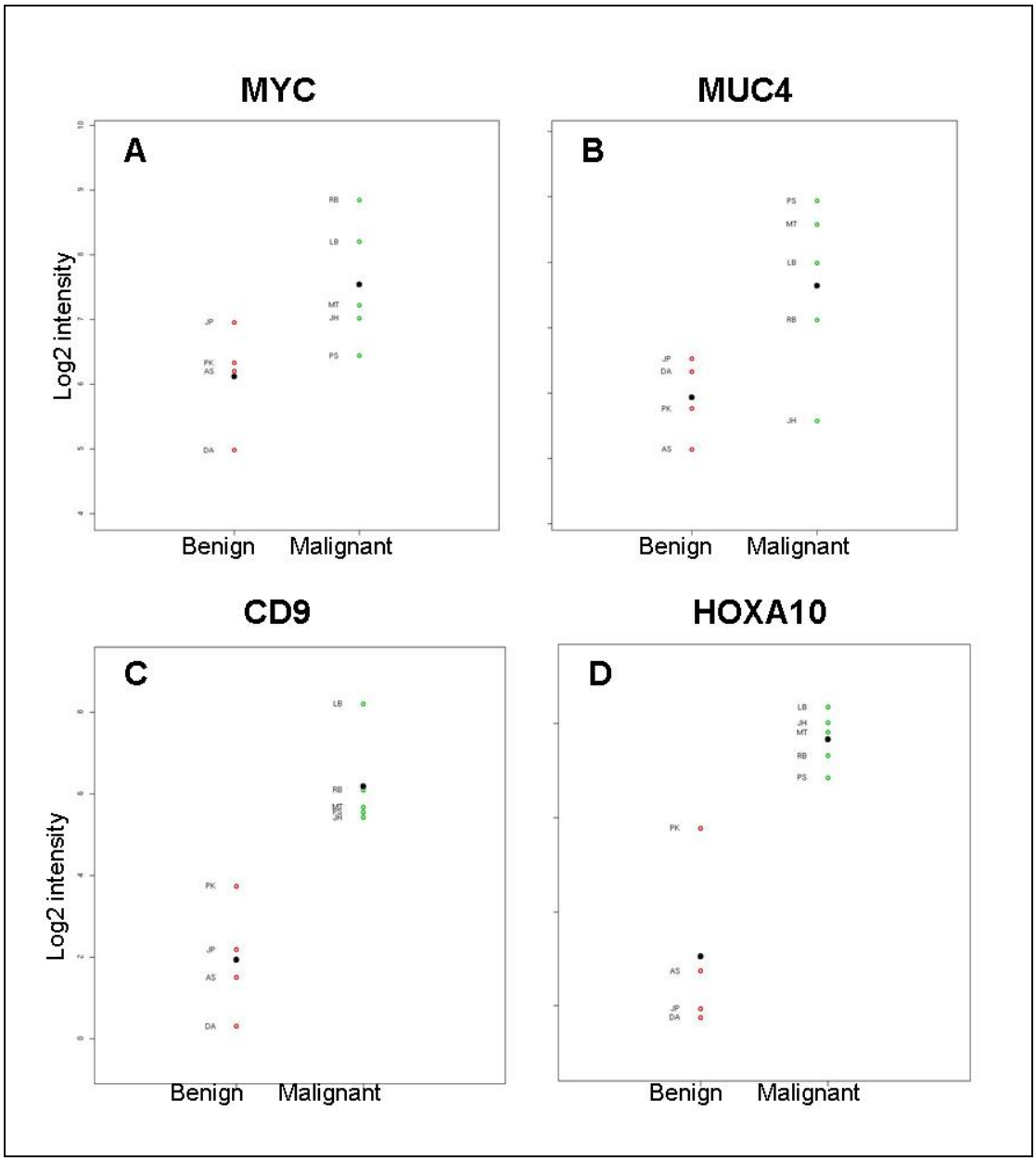


Figure 41 Examples of gene expression plots derived from Affymetrix U133 Plus 2.0 microarray analysis of biliary RNA from patients with benign (n=4) and malignant (n=6) disease, showing significantly different ($p<0.05$) expression of: A&B) two genes already reported to be up-regulated in BTC (MYC & MUC4), and C&D) two potential

novel biomarkers with no overlap between benign and malignant disease (CD9 & HOXA10)

5.4.2.5 Genes already identified as biomarkers in other cancers

It is possible that genes listed in our data have already been identified as biomarkers in other cancers but not yet assessed in hepatobiliary disease. Depending on their accuracy in other cancers, they may represent suitable candidate genes for testing in BTC.

5.4.2.6 Biological plausibility

Many genes are immediately recognised as genes important in cancer biology such as control of DNA replication, angiogenesis, cell motility etc. Others are less obvious as genes of interest but literature and database searches often identify clear cancer links generating further interest in these genes for biomarker selection. In addition, genes coding for nuclear proteins are less attractive for bile or serum protein biomarker work as they are less likely than extracellular matrix proteins to be released from the cell into the clinical samples of interest.

A tool used to make this aspect of the work simpler is software that groups genes into functions, cellular location, cellular pathways etc. This allows identification of genes encoding proteins that may be more likely to be released into the circulation for biomarker use, e.g. extracellular matrix proteins. The software used for this project was the GeneGo MetaCore

software (GeneGo, USA) in use by colleagues at CRUK (Dr Charles Swanton). The data listed in Figure 42 are lists of general cellular functions or pathways most commonly altered in our microarray data without the specific genes listed. This may be helpful in order to plan future translational in vitro experiments using BTC cell lines to further investigate pathways of interest in BTC. Table 14 below is an abbreviated list of specific genes upregulated in BTC as identified by cellular location using this MetaCore software. We concentrated on genes related to secretory pathways, extracellular matrix, cell membranes and adhesion proteins in order to concentrate on genes or protein products more likely to be idenitifable for biomarker development using simpler biologic fluids such as serum or bile. The location of the protein product of the genes of interest was one of the factors used to draw up the later shortlist of genes for further testing by qPCR.

The MetaCore software also generates maps for cellular pathways of interest with genes identified as being up- or down-regulated marked on the map. Examples of the gene grouping and pathways generated for common cell adhesion and EGFR signialling pathways are shown in Figures 42 . The cell adhesion pathway was chosen for similar reasons as outlined above- *i.e* we hypothesised that cell membrane or secreted proteins are more likely to be identifiable in serum samples than intracellular proteins for biomarker development. The EGFR pathway is shown as this pathway has previously been reported as activated and potentially implicated in the pathogenesis of BTC as well as being a potential pathway for therapeutic interventions using EGFR inhibitors (Kiguchi, Ruffino et al. 2005; Wise, Pilanthananond et al.

2008; Wang, Maass et al. 2009; Gruenberger, Schueller et al. 2010). Of particular interest for future biological investigation might be whether multiple genes with direct links to one another are identified as altered within the same pathway rather than random, potentially unconected alterations within pathways which may conceivable be due to chance alone. However, to generate such information would likely require larger sample sizes and addition of down-regulated genes in the same pathway maps as upregulation of one gene may result in downregulation of genes downstream in the pathway via other homeostatic mechanisms. Our pathway maps only show upregulated genes (>2 fold change) with a view to potential biomarker work but could in the future be revised and re-formated so show additional down-regulate genes.

Figure 42 MetaCore software groupings of commonest pathways and gene locations of up-regulated genes in BTC

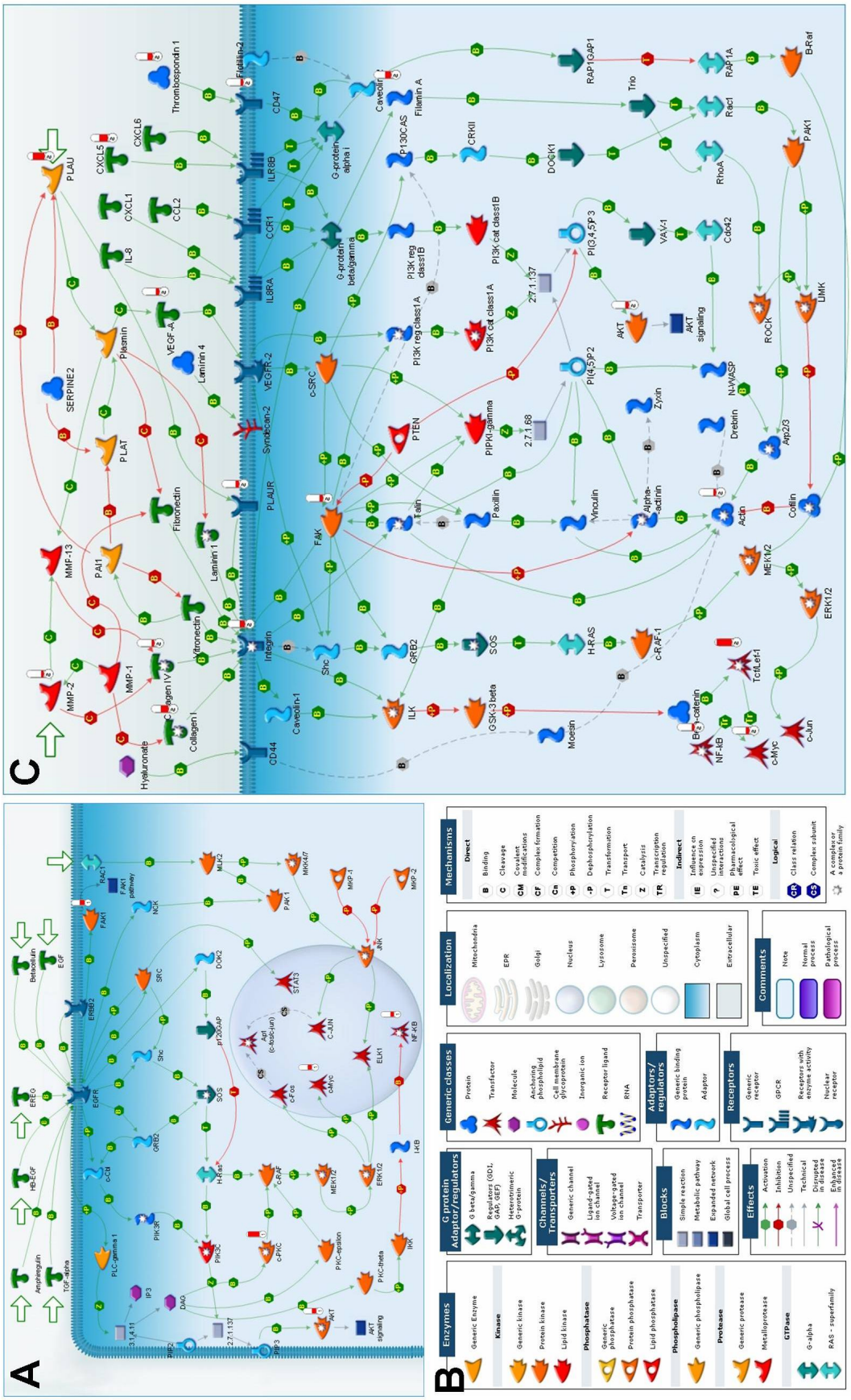


Figure 43 MetaCore software-generated EGFR and cell adhesion pathways with up-regulated genes marked (white mark with red band)

Table 14 Genes up-regulated in BTC categorised by cellular location (excludes nuclear location)

ECM	proteinaceous ECM	Anchoring collagen	Basement membrane	Cell surface	Extracellular region 1	Extracellular region 2	Secretory granule membrane
COL27A1	COL27A1	COL16A1	Osteonectin	PKHD1	HS6ST2	GALNT2	ICA69
Osteonectin	Osteonectin	Endostatin	LAMC2	MFGE8	Gpr172b	Endostatin	Rab-27A
COL4A2	COL4A2	COL6A3	COL4A2	OATP-D	MSP	MXRA5	
BPAG2	BPAG2	COL9A2	LAMA5	ITGAM	-sensitive amino	Nectin-4	
LAMA5	LAMA5		CFDP1	IL17BR	NAP-2	Oct-3/4	
COL1A1	PGAR		ACES	CD28	Ephrin-A	CD9	
Endostatin	CSPG4 (NG2)		VEGF-A	IRAK2	OSCAR	Thrombospondin 1	
VEGF-A	COL1A1		Endostatin	CSPG4 (NG2)	IL18RAP	SC1	
LAMC2	TGM2		CD43	ITGAX	CLMP	Activin beta A	
Oct-3/4	Endostatin			Calreticulin	Beta-2-M	PGES	
SC1	VEGF-A			CX3CL1	COL1A1	COL9A2	
COL9A2	COCH			HNRPU (SAF-A	VEGF-A	Activin	
COL6A3	Calreticulin			PLAUR (uPAR)	KRTCAP2	PAPLN	
Tapasin	LAMC2			ACES	Ephrin-A1	Tapasin	
COL16A1	Oct-3/4			IL-2R gamma cha	LONRF3	ACES	
Mucin 5AC	SC1			CD22	LIF	CEACAM3	
ACES	COL6A3			ICAM1	LISCH7	PLAU (UPA)	
Lumican	COL9A2			CD43	Mucin 4	lycan-like sulfated glycoprotein	
CFDP1	Mucin 4				PILRA	LILRA5	
CD43	Cyr61				Cyr61	Ppif	
	PAPLN				COL16A1	Granzyme B	
	Mucin 5AC				MAG	receptor epsilon subunit	
	COL16A1				DHRS13	PC7	
	Tapasin				CFDP1	STS	
	ACES				CARD5	Alpha-defensin	
	MMP-2				IRAKM	MFGE8	
	CFDP1				C1orf56	Osteonectin	
	Lumican				ATR/TEM8	COL27A1	
papilin, proteoglycan-like sulfated glycoprotein					SERPINA3 (ACT	IDS	
	CD43				FUT6	CEACAM1	
					P-cadherin	IL17BR	
					Semaphorin 4C	Fetuin-A	
					PAP39	Fibrosin	
					LAMA5	Alpha-defensin 3	
					PGAR	C1r	
					Ceruloplasmin	Neuromedin U	
					TNFAIP2	GRO-2	
					CSPG4 (NG2)	CNIH3	
					GZMH	SRPUL	
					ENA-78	Collagen IV	
					TGM2	LAMC2	
					NPEPPS	SIAT4B	
					COCH	c gamma RII alpha	
					Calreticulin	NOTCH3	
					CRISPLD2	PDK1	
					CX3CL1	CD47	
					PLOD3	Lumican	
					PLAUR (uPAR)	COPA	
					TRSC 1	(Tripeptidyl-peptidase I)	
					ITIH5	PCDHB16	
					COL6A3	CD22	
					CELSR1	CD43	
					Furin	BPAG2	
					GLVR1	COL4A2	
					Mucin 5AC	ATP6V0A2	
					CD109	Alpha-defensin 1	
					MMP-2	Activin A	
					Calgizzarin		

5.4.3 Shortlist of genes for further validation using TaqMan arrays

The shortlist drawn up of candidate genes for validation using qPCR on the TaqMan arrays (n=128) contained a selection of the 1,140 up-regulated and 1,001 down-regulated genes identified by microarray (Table 15). Examples of reasons for inclusion are given in the following categories:

5.4.3.1 Ubiquitous constitutively expressed genes for normalisation and quantification of qPCR data

The genes chosen for this use were Glyceraldehyde 3-phosphate dehydrogenase [GAPDH], 18S ribosomal RNA, [18SRNA], B-actin [ACTB] and beta glucuronidase [GUSB].

5.4.3.2 Biomarkers already reported to be related to BTC

These included mucin 4 [MUC4], mucin 5AC [MUC5AC], myc oncogene [c-MYC], minichromosome maintenance protein 4 [MCM4], mitogen-activated kinase 1[MAPK1] and HOX genes.

5.4.3.3 Genes with very high increased fold change in cancers

These included paraneoplastic antigen A2 [PNMA2] (x39), matrix remodelling associated 5 [MXRA5] (x24), inter alpha inhibitor 5 [ITIH5] (x20), serpin

peptidase inhibitor clade A [SERPINA3] (x40), collagen 17α1 [BPAG2] (x16) and CD9 (x19).

5.4.3.4 Genes with no overlap expression in cancers compared to benign as seen with the gene plot graphs

These included CD9, integrin B8 [ITGB8], vascular endothelial growth factor A [VEGFA], carcinoembryonic antigen cell adhesion molecule 1 [CEACAM1], inter alpha inhibitor 5 [ITIH5], platelet endothelial aggregation receptor 1 [PEAR1], aspartate beta hydroxylase [ASPHD1], homeobox A10 and B6 [HOXA10 & HOXB6] and craniofacial development protein1 [CFDP1].

5.4.3.5 Genes identified as tumour markers in other cancers

Examples include matrix remodelling associated 5 [MXRA5], paraneoplastic antigen A2 [PNMA2] in colorectal cancer, [RAD51] in pancreatic cancer, anthrax receptor toxin 1 [ANTXR1] in colorectal cancer, mitogen-activated kinase 1 [MAPK1] in many adenocarcinomas, IL18 receptor accessory protein [IL18RAP] in adenocarcinomas and carcinoembryonic antigen cell adhesion molecule 1 [CEACAM1] in many other adenocarcinomas.

5.4.3.6 Genes related to tumour biology

Examples include the homeobox genes that control cell proliferation and development, notch signalling genes, protein kinases [MAPK1, PRKCB1, PTK2, TNFAIP2, CSPG4], minichromosome maintenance proteins [MCM4],

angiogenesis [VEGFA], matrix remodelling genes [MXRA5, MMP2, CSPG4], oncogenes [cMYC, KRAS, PVT1], and genes involved in cell adhesion and motility [CEACAM1, ITGB8, MUC4, CD9] .

5.4.3.7 Down-regulated genes

Only a small number of genes were included in this group: mastermind like 3 [MAML3], Sel-1 suppressor of lin 12 [SEL1L], tumour suppressor homolog 3 [FAT3] and KRAS.

Table 15. Genes identified in microarray data for consideration of further validation using qPCR. Most genes in black were selected for further investigation; those in grey were not further assessed. In the notes column, overlap in gene expression plots refers to the results presented in the microarray gene plot graphs.

Gene	Name	Fold Change	P	Function	Notes
Up-regulated genes					
MUC4	Mucin 4	3.2	0.025	Glycoprotein	Extracellular space, reported in BTC, some overlap in gene plots. TaqMan amplicon length 55bp
MUC5AC	Mucin 5AC	3.6	0.033	Glycoprotein	Extracellular space, reported in BTC, some overlap in gene plots
CD9	CD9	19	0.0001	Cell surface protein	Cell adhesion & motility, no overlap in gene expression plots. TaqMan amplicon 72
LUM	Lumican	11	0.001	ECM protein	Collagen binding, visual perception, No overlap in gene expression plots. TaqMan amplicon 107
MAML2	Mastermind like -2	3.5	0.00085	Notch signalling	Notch signalling, nuclear, control of transcription, no overlap in gene expression plots. TaqMan amplicon 84
PMPFIBP1	PTPRF interacting protein binding protein 1	2.6	0.010	Cell adhesion	S100A4 interaction, tumour invasiveness, axonal growth, little overlap in gene expression plots
ITGB8	Integrin B8	2.7	0.0011	Cell adhesion	Cell adhesion, no overlap in gene expression plots. TaqMan amplicon 64
VEGFA	Vascular endothelial growth factor A	2.5	0.001	VEGF, cell adhesion	VEGF, cell adhesion, cancer, no overlap in gene plots. TaqMan amplicon 63
PNMA2	Paraneoplastic antigen a2	39	0.002	Nucleus	Nuclear protein, function not clear. Found in serum in testicular and neuronal cancer, some overlap in gene plots. TaqMan amplicon 60
CEACAM1	Carcinoembryonic antigen cell adhesion molecule 1 (biliary glycoprotein)	3.5	0.0025	Cell adhesion	Cell adhesion, angiogenesis, cell motility. No overlap in gene plots. TaqMan amplicon 78
MXRA5	Matrix remodelling associated 5	24	0.007	Cell adhesion	Extracellular binding, overlap in gene expression plots, up regulated in colon cancer, function unknown. TaqMan amplicon 88
THBS1	Thrombospondin 1	4.6	0.0084	Cell motility, cell adhesion	Cell adhesion, protein binding, TGF B signalling, Little expression overlap,
HOXA10	Homeobox A10	24	0.00011	DNA transcription factor	Nuclear, DNA transcription, no overlap in gene expression plots. TaqMan amplicon 52
HOXB6	Homeobox B6	20	0.00023	Development and cell division	Nuclear transcription, no overlap in gene plots TaqMan Amplicon 73
ITIH5	Inter alpha (globulin) inhibitor H5	20	0.0008	Hyaluronan metabolism	Serine endopeptidase inhibitor activity, No overlap in gene plots, ABI amplicon 67
PEAR1	Platelet endothelial aggregation receptor 1	10.3	0.0017	Platelet aggregation	Membrane protein, No overlap in gene plots. TaqMan amplicon 66
RAPGEF3	Rap guanine nucleotide exchange factor 3	10.2	0.0024	Cell proliferation, GTPase, leukocyte migration	Membrane protein, no overlap in gene plots
ASPHD1	Aspartate beta hydroxylase domain 1	10	0.0033	Peptidyl amino cid modification	Endoplasmic reticulum, no overlap in gene expression plots. TaqMan amplicon 62
TGFBI/1	TGF B induced transcript 1	12.3	0.003	Cell adhesion, Cell differentiation	Wnt signalling, cell adhesion, RNA polymerase promoter, tight junctions, No overlap in gene plots. TaqMan amplicon 70
TM4SF18	Transmembrane 4 L six family member 18	10.9	0.0042	Membrane protein	Membrane protein. Function unknown, little overlap in gene plots. TaqMan amplicon 70
PVT1	PVT oncogene homolog	11	0.0055	Function unknown	Function unknown. Little overlap in gene plots. TaqMan amplicon 111
LEF1	Lymphoid enhancer binding factor 1	15	0.0069	Developmental biology, RNA polymerase II promoter,	Wnt signalling, DNA binding, multiple cancer reports, little overlap in gene plots. TaqMan amplicon 60

LILRA2	Leukocyte Ig like receptor 2	11.5	0.0082	Antigen response	Cell membrane, signal transduction, antigen binding, little overlap in gene plots
PAPLN	Papilin, proteoglycan-like sulphated glycoprotein	2.8	0.020	Metallopeptidase	Proteinacious ECM , overlap in gene plots
CFDP1	Craniofacial development protein 1	5.3	0.0005	Cell adhesion	Proteinacious ECM, anti-apoptosis, no expression overlap. TaqMan amplicon 63
MMP-2	Matrix metallopeptidase 2	2.6	0.020	Matrix metallopeptidase	Proteinacious ECM, collagenase, some overlap in gene plots. TaqMan amplicon 65
COL16A1	Collagen 16	2.5	0.017	Cell adhesion	Proteinacious ECM, overlap in gene plots
SPARC	Osteonectin (Secreted protein, acidic rich)	7.1	0.004	Kinase signalling	Proteinacious ECM, collagen binding, kinase signalling, ossification, little overlap in gene plots. TaqMan amplicon 76
Cyr61	Cystein rich angiogenic inducer 61	3.2	0.03	Cell growth & adhesion	Proteinacious ECM, cell proliferation, overlap in gene plots
COL6A3	Collagen 6 a3	8.8	0.003	Cell adhesion	Proteinacious ECM, cell communication, little overlap in gene plots
TCF19 (SC1)	Transcription factor 19	2.2	0.023	Transcription, cell proliferation	Proteinacious ECM. significant overlap in gene plots
POU5F1 (Oct-3/4)	POU class 5 homeobox 1	3.3	0.0018	Regulation of transcription	Proteinacious ECM, no overlap in gene expression plots, reports in cancers, Stem cell marker? TaqMan amplicon 77
CALR	Calreticulin	2	0.026	apoptosis	Proteinacious ECM, gene expression overlap in gene plots Antigen presentation, control of proliferation and apoptosis
COCH	Coagulation factor C	6.1	0.012	Sensory perception of sound	Proteinacious ECM, function unknown, some overlap in gene plots
TGM2	Transglutaminase 2	3	0.016	Cell adhesion, apoptosis	Proteinacious ECM, NF kappa B signalling, blood vessel remodelling, some overlap in gene plots
LAMC2	Laminin gamma 2	2.5	0.005	Cell adhesion	Proteinacious ECM , ECM signalling, cell adhesion, development, no overlap in gene plots. TaqMan amplicon 79
LAMA5	Laminin alpha 5	2.33	0.021	Cell adhesion, development	ECM communication and cell adhesion, Proteinacious ECM, overlap in gene plots
COL18A1	Endostatin	2.8	0.014	Anti-angiogenesis, cell adhesion	Endogenous inhibitor of angiogenesis. , development, Proteinacious ECM, overlap in gene plots
ICA1	Islet cell autoantigen 1	2.1	0.030	Golgi membrane	Secretary granule, overlap in gene plots
RAB27A	Ras oncogene family	2.5	0.019	RAS oncogene	GTPase activity, protein transport, Secretory granule, little overlap in gene plots. TaqMan amplicon 121
TAPBP	Tapasin	2.14	0.037	Antigen processing,	Cell surface, retrograde endosome transport, protein complex assembly, little ovelap in gene plots
MFGE8	Milk fat globule EGF factor 8	7.5	0.009	Cell adhesion	Cell surface, some overlap in gene plots
PKHD1	Polycystic kidney and hepatic disease 1	2	00.025	Cell adhesion	Cell surface. No overlap in gene plots, negative effects cell motility. Haemostasis
ITGAM	Integrin alpha m, complement subunit	2.9	0.019	Cell adhesion, cell motility	Cell surface, effects on actin cytoskeleton, cell motility, some overlap in gene plots
ICAM1	Intercellular cell adhesion molecule 1	3.6	0.037	Cell adhesion	Cell surface, cell adhesion, NK cell mediated killing, trans endothelial cell migration, wide overlap
CD22		8.2	0.02	Cell adhesion	Cell surface, plasma membrane, no overlap expression, mostly B lymphocytes
IL2RG	IL-2R gamma chain	10	0.024	Cell membrane signalling	Cell surface, jak-stat signalling, immune responses, wide overlap in expression
PLAUR (uPAR)	Plasminogen activator urokinase receptor	3.1	0.009	Cell motility, complement and coagulation cascades	Cell surface, Cell motility, complement and coagulation cascades, some spread within groups in gene plots
CD28	CD28	3.3	0.038	T cell activation	Cell surface, T cell proliferation, anti-apoptosis, cytokine synthesis, wide overlap in expression plots
IL17RB	IL17 receptor B	2.3	0.004	Cytokine response	Cell surface, cytokine to cytokine interactions, activation of immunocytes, little overlap in gene plots
IRAK2	IL1 receptor kinase 2	3.4	0.006	apoptosis	Cell surface, kinase signalling, NFKB

					pathway, apoptosis, inflammatory response, little overlap in gene plots
CSPG4 (NG2)	Chondroitin sulphate proteoglycan	8.5	0.018	Angiogenesis	Cell surface, Angiogenesis, cell motility, cell differentiation, development, tissue remodelling, protein tyrosine kinase, Some overlap in gene plots. TaqMan amplicon 113
ITGAX	Integrin alpha x (complement component)	4.8	0.002	Cell adhesion	Cell surface, cell adhesion, integrin mediated signalling, No overlap in gene plots
CX3CL1	Chemokine ligand 1	5.4	0.022	Cell adhesion	Cell surface, leukocyte activation and adhesion, cytokine receptor interactions, some overlap in gene plots
HNRNP U (SAF-A)	Heterogeneous ribonuclear protein	2.03	0.021	mRNA processing	Cell surface, significant overlap in gene expression plots
PRKCB1	Protein kinase C beta 1	8.3	0.014	Protein kinase phosphorylation	Protein kinase and VEGF signalling, EGFR pathway, some overlap in gene plots. TaqMan amplicon 68
PTK2	Protein tyrosine kinase 2	3	0.03	ERBB2 binding and activation	Cell membrane, intracellular, EGFR, no overlap in gene plots benign/CC. TaqMan amplicon 68
cMYC	Myc oncogene	2.7	0.028	Cell proliferation, cell cycle arrest,	Regulation of transcription, jak-stat signalling, wnt signalling, TGFb signalling, cancers++, some overlap of gene plots. TaqMan amplicon 65
MFGE8	milk fat globulin EGF factor 8	7.5	0.009	Cell adhesion	Cell adhesion, some overlap in gene expression plots
LIF	Leukaemia inhibitory factor	4.7	0.0012	Jak-stat signalling	ECM, PK pathways, cell proliferation, no overlap in gene plots, biomarker in other cancers. TaqMan amplicon 66
UBE2D3 (MAG)	Ubiquitin-conjugating enzyme E2D 3	2.1	0.026	Ubiquitin cycle	Ubiquitin mediated proteolysis, some overlap in gene plots
Ephrin-A1	Ephrin A1	2.4	0.03	Axon guidance	Cell membrane, Cell signalling, significant overlap in gene plots
IRAKM	Interleukin receptor kinase 3	5	0.009	apoptosis	Cytokine and chemokine signalling, apoptosis, some overlap in gene plots
ANTXR1	Anthrax toxin receptor 1	3.5	0.012	Membrane receptor	Tumour specific endothelial marker in colorectal cancer, some overlap in gene plots
SERPIN A3 (ACT)	Serpin peptidase inhibitor clade A	37.8	0.012	Acute phase response	DNA binding, inflammation, regulation of lipid metabolism, some overlap of gene plots. TaqMan amplicon 70
HS6ST2	Heparin sulphate sulphotransferase 2	3.1	0.029	Heparin sulphate biosynthesis	Cell membrane, significant overlap in gene plots
Gpr172b	G protein receptor 172b	2.7	0.002	Membrane receptor	Cell membrane, no overlap in plots
PPBP (NAP-2)	Pro-platelet basic protein	9.3	0.010	Chemotaxis	Cell proliferation, ECM, some overlap in gene plots
OSCAR	Osteoclast associated Ig-like receptor	2.4	0.002	Plasma membrane receptor	Plasma membrane receptor, osteoclast differentiation, presentation to dendritic cells, no overlap in gene plots
IL18RAP	IL18 receptor accessory protein	6.6	0.0004	Cell membrane receptor	Inflammatory response, no overlap in gene plots, expressed in some epithelial cancers
ASAM (CLMP)	Adipocyte specific adhesion molecule	3.1	0.026	Tight junction protein	Tight junction, significant overlap of gene plots
B2M	Beta-2-microglobulin	2.4	0.007	Membrane receptor	Antigen processing and presentation, little overlap in gene plots
KRTCAP 2	Keratinocyte associated protein 2	2.1	0.02	Cell membrane protein	Oligosaccharyl transferase complex, some overlap
CELSR1	Cadherin, EGF LAG seven pass G type receptor 1	5.2	0.004	Cell membrane protein	G protein activity, cellular polarity, neural tube closure, development, protein dimerisation, no overlap in gene plots. TaqMan amplicon 99
SPRY4	Sprouty homolog 4	3.5	0.004	Cell membrane	Development, signal transduction, jak-stat signalling, some overlap in gene plots
CDH3	Cadherin 3 type 1, placental (p-cadherin)	6	0.012	Cell adhesion	Cell adhesion, some overlap in gene expression plots
NPEPPS	Aminopeptidase puromycin sensitive	2.1	0.022	proteolysis	Aminopeptidase, metalloproteinase, some overlap in gene expression plots
GZMH	Granzyme H (cathepsin G like 2)	6.6	0.025	cytoplasm	Apoptosis, serine endopeptidase activity, some overlap in gene plots.
TNFAIP2	TNF, alpha-induced protein-2	2.8	0.018	Extracellular space	Angiogenesis, development, cell differentiation, Little overlap in gene plots
ANGPTL	Angiopoeitin like 4	4.2	0.033	Extracellular	Angiogenesis, cell differentiation,

4				space	development, response to hypoxia, PPAR pathway, some overlap in gene plots
PRPSAP 1 (PAP39)	Phosphatidyl pyrophosphate synthetase-associated 1	2.3	0.019	Nuclear protein	Nucleic acid metabolism/ synthesis, enzyme inhibitor, significant overlap in gene plots
PGES	Prostaglandin E synthase	5.3	0.028	Arachidonic acid metabolism	Significant expression overlap in gene plots
INHBA	Activin beta A, inhibin beta A	3.86	0.031	Cell proliferation	Cell proliferation and differentiation, significant overlap in gene expression plots
SC1	Transcription factor 19	2.22	0.020	Cell proliferation	Regulation of transcription from RNA, overlap in gene plots
ATP6V0 A2	Lysosomal H transporting subunit V0 a2	10	0.0010	ATPase, proton pump	Membrane protein, Proton pump, epithelial cell signalling, No overlap in gene expression plots. TaqMan amplicon 58
PCDHB1 6	Protocadherin beta 16	2.4	0.016	Cell adhesion	Cell adhesion, some overlap in gene plots
CLN2 (TPP1)	Tripeptidyl-peptidase I	2	0.01		Tripeptidyl-peptidase I
S100A11	Calgizzarin	2.6	0.02	S100A11 calcium binding protein	Signal transduction
SLC20A 1 (GLVR1)	Solute carrier family 20 (phosphate transporter) member 1	2.4	0.019	Plasma membrane	Phosphate metabolism, positive regulation of I kappa B pathway, some overlap in gene plots
FURIN	Furin (paired basic amino acid cleaving enzyme)	2.9	0.033	Plasma membrane	Golgi peptise synthesis and transport, furin activity, calcium binding, cell-cell signalling, overlap in gen plots
PLOD3	Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3	2.7	0.009	Endoplasmic reticulum, protein modification	ER protein modification, procollagen alteration, metal ion binding, virtually no overlap in gene plots
STS	Steroid sulfatase (microsomal isozyme S	2.1	0.020	Plasma membrane, Androgen and oestrogen metabolism	ER, microsomes, lipid, androgen, oestrogen metabolism, calcium binding and hydrolase activity, some overlap in gene plots
GZMB	Granzyme B	2.3	0.03	cytoplasm	NK cell killing, proteolysis, apoptosis, cleavage of lysine, some overlap in gene plots
GABRE	GABA-A receptor epsilon subunit	3.4	0.003	GABA signalling	GABA signalling, very little overlap in gene plots
PCSK7	Proprotein convertase subtilisin /kexin type 7	2.3	0.010	Cell membrane	Proteolysis, ,golgi, no overlap
Ppif	Peptidylproolyl isomerase F (cyclophilin F)	2.6	0.010	mitochondria	Protein folding, iosomerase activity, some overlap in gene plots
CEACA M3	Carcinoembryonic antigen cell adhesion molecule 3	3.1	0.029	Cell membrane	Cell adhesion, signifiacant overlap
SEMA4C	Semaphorin 4C	2	0.018	Cell membrane receptor	Cell differentiation, development, axon guidance, no overlap in gene plots
COPA	Coatomer protein complex subunit a	4.7	0.010	ER, Golgi, protein folding	Protein folding, vesicles, ER, Golgi, some overlap in gene plots
PDK1	Pyruvate dehydrogenase kinase 1	2	0.014	T cell receptor signalling	Carbohydrate & glucose metabolism, mitochondrial protein, little overlap in gene plots
NOTCH3	Notch homolog 3	6	0.022	Plasma membrane, Notch signalling	Development, regulation of transcription, cell differentiation, some overlap in gene plots. TaqMan amplicon 87
CNIH3	Cornichon homolog 3	2	0.022	Intracellular signalling	Membrane protein, some overlap in gene plots
CXCL2 (GRO-2)	Chemokine (CXC motif) ligand 2 or Gro 2 oncogene	4.7	0.028	chemotaxis	Chemokine cytokine interaction, g protein signalling, extracellular space, some overlap in gene plots
NMU	Neuromedin U	3.6	0.024	Smooth muscle contraction	Extracellular space, neuropeptide signalling, muscle contraction, some overlap in gene plots
C1r	Complement component 1 r subcomponent	2.2	0.018	complement	Complement and coagulation cascade, ECM matrix, little overlap in gene plots
FBRS	Fibrosin	2.4	0.039	?	? significant overlap in gene plots

IDS	Iduroniate 2 sulfatase	2.3	0.02	Glycosaminoglycan degradation	Hunters syndrome, ER location, overlap in gene plots
COL4A2	Collagen type IV, alpha 2	2.4	0.014	Cell adhesion	Proteinacious ECM, cell adhesion and communication, ECM receptor interaction, no overlap in gene plots
COL1A1	Collagen type 1 a1	7.2	0.00049	Cell adhesion	Proteinacious ECM, no overlap in gene plots. TaqMan amplicon 66
COL9A2	Collagen type IX, alpha 2	2.1	0.025	Skeletal development	Some overlap in gene plots
COL16A1	Collagen XVI alpha1	2.4	0.017	Cell adhesion	ECM protein, integrin signalling, overlap in gene plots
BPAG2 (COL17A1)	Collagen 17 a1	16	0.019	Cell adhesion	Proteinacious ECM, cell adhesion, cell communication, wide variation in cancer samples. TaqMan amplicon 64
COL27A1	Collagen 27, alpha 1	2.2	0.024	Cell adhesion	Proteinacious ECM, overlap in gene plots
CTNND1	Catenin (cadherin associated protein) delta 1	3.4	0.005	Transcription, cell adhesion, adherens junction	Wnt signalling, cell-cell adhesion, regulation of transcription, leukocyte trans endothelial migration, almost no overlap in gene plots
MCM4	Minichromosome maintenance complex 4	2.2	0.03	Control of DNA replication	DNA replication, nuclear protein, overlap in plots, identified in BTC cell line microarray data also, TaqMan amplicon 90
MAPK1	Mitogen-activated protein kinase 1	2.35	0.010	MAPK signalling, apoptosis, cell cycle	MAPK signalling, apoptosis, cell cycle, chemotaxis, response to DNA damage, VEGF signalling, TGF signalling, reports in cancer including CC, pancreas. No overlap in gene plots. TaqMan amplicon 74
RAD51	RAD51 homolog	4.3	0.03	DNA repair	Some overlap, data in pancreatic cancer. TaqMan amplicon 58
NCSTN	Nicastrin	3.4	0.004	proteolysis	Notch signalling, no overlap in gene plots. TaqMan amplicon 69
STAT1	Signal transducer and activation of transcription 1	2.5	0.002	DNA transcription	JAK-STAT signalling, nuclear protein, no overlap in gene plots. TaqMan amplicon length 67
TRIB2	Tribbles homolog 2	4.3	0.001	MAPK	MAPK, protein phosphorylation, cytoplasm, no overlap in gene plots. TaqMan amplicon length 60
LYST	Lysosomal trafficking regulator	3.7	0.003	Endosome transport	Microtubule function, endosomes, cellular defence responses, no overlap in gene plots. Gene defect of Chediak Higashi disease
HDAC9	Histone deacetylase complex 9	2.3	0.019	DNA transcription	DNA replication, development, inflammatory response, some overlap in gene plots, HDAC2 key in HCC
Down-regulated genes					
FAT3	Tumour suppressor homolog 3	-8.8	0.014	Cell adhesion	Calcium binding, some overlap in gene plots
SEL1L	Sel-1 suppressor of lin12	-7.5	0.011	Notch signalling	Membrane and ER protein, notch signalling, wide spread in cancer but little overlap in gene plots
KRAS	KRAS	-2.5	0.013	MAPK signalling	GTPase, development, NK cell killing, cancer, some overlap in gene plots. TaqMan amplicon 109
MAML3	Mastermind like 3	-8.5	0.007	Notch signalling	Large variation on cancers on gene plots. TaqMan amplicon 104

5.5 Conclusions

In this chapter we demonstrate that it is possible to gain useful whole genome gene expression data from RNA isolated from biliary brushings. Evaluation of the genes with significant fold change in gene expression identifies many of the genes already noted to be altered in BTC as well as many that may be useful for the identification of novel biomarkers for BTC. Confirmation of gene expression profiles using an alternative method such as qPCR is required prior to testing protein expression of the candidate genes.

Our data also justify further investigation using a larger sample size in order to identify a gene signature diagnostic of cancer in biliary brushings or to use similar methodology in order to investigate the biology of disease progression and development of cholangiocarcinoma in patients with PSC.

5.6 Discussion

Despite the encouraging results from our microarray data, the possibility that results represent a high level of false positivity remains. The first broad cause is methodological. The degraded nature of the RNA, small samples size and very large number of genes analysed has the potential for introducing significant variation by chance alone. RNA degradation is outside of our control. Sample size is dependent on availability of samples and cost, which are the primary limiting factors. Statistical error is controlled for as best as possible using statistical methods designed to reduce this error to minimal

levels. Correction for small sample size and multiple hypothesis testing were used in our methodology.

The second significant source of potential error is the presence of confounding mRNA from other cell types including leukocytes, red cells and fibroblasts. Other studies have compared surgical resection tissue to normal liver which introduces significant bias in the predominance of hepatocytes in the control arm and cancer cells and stromal tissue in the other. Our samples were similar clinical samples from both malignant and benign groups and should reduce this error by having more similar quantities of epithelial and fibroblastic cells in each group. A major source of potential confounding material is leukocytes in obstructed, non sterile bile ducts. Comparison of cytology and clinical details (including evidence of overt cholangitis), does not suggest a difference between benign and malignant groups. In addition, leukocyte specific genes were not differentially expressed in either group providing strong evidence against a significant leukocyte confounding effect. Also, our experiments validating the RNA methodology in chapter 4 show that RNA is predominantly epithelial cell in origin and that leukocytes have a relatively small contribution to the overall mRNA pool. All samples must have included RNA from non epithelial cell sources but the assumption is that similar amounts of such RNA are found in both groups and the effect is therefore cancelled out during data analysis.

Analysis of our data has shown a number of genes noted to be important in cancer biology and in particular, the biology of BTC. This is also very

encouraging and strongly suggests that the data is accurate. Many of these genes are genes that are easily identifiable as having functions related to cancer biology such as cell replication, adhesion and motility. However, such genes are likely to be non specific for BTC although may be useful biomarkers of cancers in general. It is also likely that the specificity of such genes will be reduced by the presence of other actively replicating cells such as leukocytes in infection or inflammatory disease. It may be that the most useful protein biomarkers will be derived from unusual or unexpected proteins found to be up-regulated and such genes should therefore be included in the search, even if they appear bizarre. However, other, more predictable genes, may be useful tests when combined as customised gene expression arrays and are therefore included in the next stage of the project.

Chapter 6.

6 Validation of gene expression profile using real time quantitative PCR

6.1 Background

Microarray analysis generates huge amounts of data that are too large to be analysed individually to screen for technical or reading errors. Also, the sheer volume of data carries the potential for a significant number of false positive results, even after correction for multiple hypothesis testing. As a result, using a significance value of 5%, we would expect at least 57 of the 1140 up-regulated genes to be related to chance and not a feature of BTC. False positive results may be minimised by larger samples sizes and using a stricter p value cut-off of <1% or 0.1%, but require further testing using alternative methods of measuring gene expression.

Real time quantitative PCR (qPCR) is a well established technique used to validate patterns of gene expression in microarray studies. The methods used for qPCR are accurate and reliable and represent an alternative method to microarray for quantification and validating gene expression. However, qPCR has a number of disadvantages including:

1. It is relatively time consuming to assess multiple genes in multiple samples, despite advances in commercial kits and data analysis software.
2. Relatively large quantities of input RNA are required. Each standard qPCR reaction uses approximately 5-10ng of input RNA converted 1:1 to cDNA. When using duplicates and assessing multiple genes, low RNA yield materials such as biliary brushings can only be assessed for a limited number of genes (approximately 30 for our RNA samples).
3. Large numbers of separate experiments become costly.
4. Each PCR primer set used requires testing for specificity of target mRNA, especially when using the SYBR Green method or using non commercially available primer sequences.
5. Primer sets may amplify genomic DNA in addition to mRNA resulting in an over-estimate of gene expression and inaccurate results. This can be minimised by confirming absence of genomic DNA in the samples using reverse transcriptase negative (RT-) controls that are tested separately, or by using primer pairs designed to span two separate exons and avoid non specific genomic intron DNA amplification.

6.1.1 Part 1: The use of TaqMan Arrays for multiple gene qPCR

In order to overcome some of difficulties of individual gene qPCR, one can use commercially available customised qPCR arrays that assess multiple gene expression by Taq polymerase qPCR from a relatively small amount of input cDNA mastermix. At present, the main available assay is the TaqMan Array® (formerly TaqMan Low Density Array® [TLDA]) manufactured by Applied Biosystems Inc (Austin, TX, USA). The use of qPCR arrays is a well established technique for measurement of multiple gene expression for purposes such as validation of microarray data (Canales, Luo et al. 2006).

Each TaqMan Array (figure 44) has 384 wells per card which can be customised to include Taq polymerase qPCR primer sets for most genes of interest. In addition to target genes to be measured, the assay requires at least one (recommended 2 to 4) reference 'house keeping' genes (e.g. GAPDH, 18S rRNA, ACTB) for normalisation. As described in chapter 4, relative gene expression is calculated using the comparative threshold (ΔCt) method.

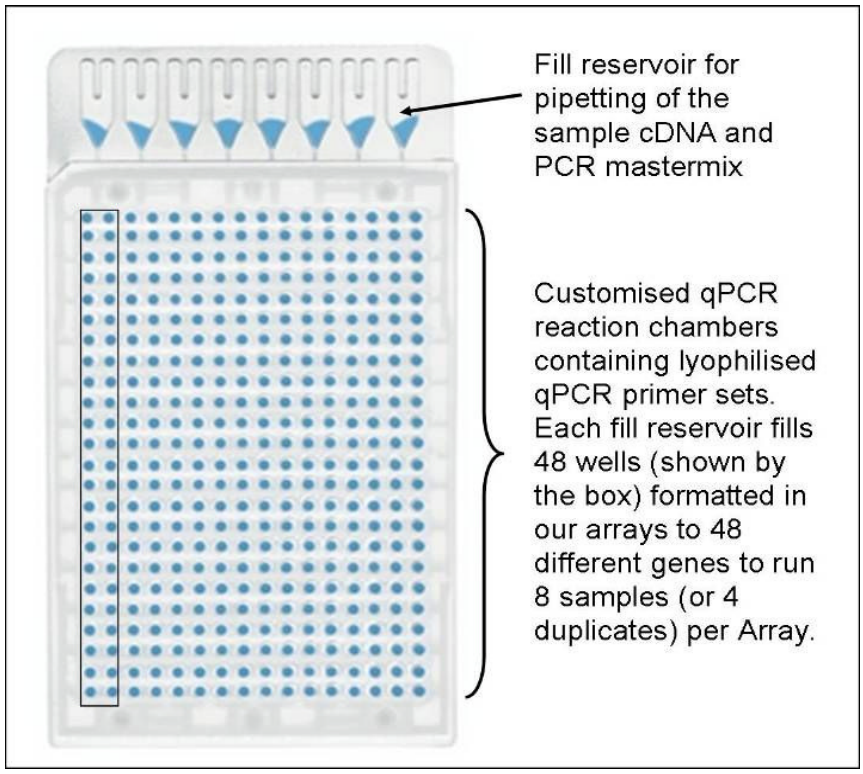


Figure 44. Format of the TaqMan Array (Applied Biosystems) used for qPCR validation. The reservoir provides the cDNA substrate and PCR mastermix that fill each of the 48 wells by centrifuging. (Original image from Applied Biosystems)

We used customised TaqMan Array format 48 which included 3 reference genes and 45 other genes for measurement of gene expression (Table 16). As discussed in chapter 5, the genes for further assessment were chosen using a variety of criteria including fold change, p value, biological plausibility, qPCR amplicon length etc.

Table 16. Genes selected for further analysis using TaqMan Array qPCR.

List of genes for qPCR validation using TaqMan arrays					
18S RNA	ITGB8	CELSR1	CFDP1	ATP6V0A2	STAT1
GAPDH	VEGFA	NCSTN	MMP-2	PRKCB1	NOTCH3
ACTB	PNMA2	ASPHD1	COL6A3	PTK2	COL1A1
MUC4	CEACAM1	TGFBI/1	SPARC	cMYC	BPAG2
MUC5AC	MXRA5	TM4SF18	TRIB2	SERPINA3	SEL1L
CD9	HOXA10	PVT1	LAMC2	LIF	MCM4
LUM	HOXB6	LEF1	RAB27A	RAD51	MAPK1
MAML2	ITIH5	POU5F1	CSPG4	MAML3	KRAS

6.2 Materials and methods

6.2.1 Clinical samples.

Clinical samples were collected at the time of ERCP as described in Chapter 4. A summary of patient details is shown in Tables 17 and 18.

Table 17. Summary of patient details for microarray sample set.

SOD- sphincter of Oddi dysfunction: CP- chronic pancreatitis

	Sample	Diagnosis	Age	Gender
Benign	**B1**	Papillary Stenosis	74	M
	B2	SOD	55	F
	B3	CBD Stricture (CP)	48	F
	B4	CBD Stricture (CP)	58	M
			Mean 59	
Malignant	**M1**	CC	50	M
	M2	CC	64	M
	M3	CC	79	F
	M4	CC	59	M
	M5	CC	79	F
			Mean 66	

Table 18. Fresh sample set used for validation of gene expression using TaqMan Array.

	Sample	Diagnosis	Age	Gender
Benign	**B5**	PSC	35	M
	B6	CBD Stricture and Stones	65	M
	B7	IgG4 Disease	69	M
	B8	SOD	38	F
	B9	IgG4 Disease	47	M
			Mean 53	
Malignant	**M6**	CC	79	M
	M7	CC	63	M
	M8	CC	60	F
	M4	CC	59	M
	M9	CC	71	F
	M10	CC	72	M
			Mean 67	

6.2.2 RNA isolation, purification and cDNA synthesis.

RNA from the biliary brushings was isolated and purified as described in chapter 4. cDNA was synthesized using the High Capacity RNA to cDNA Synthesis Kit (Applied Biosystems Inc, Austin, TX, USA) as per the manufacturers instructions. cDNA synthesis was performed in 20μl volumes from a template of 150ng total RNA using the reverse transcriptase enzyme mix for 60mins at 37^{o}C. Reverse transcriptase negative (RT-) controls were made for the majority of samples where there was potential spare RNA in order to test for genomic DNA contamination. cDNA samples were stored at -20^{o}C until further use.

6.2.3 Quantitative real time PCR.

Suitable specialist equipment was initially not available at UCL and so the TaqMan Array qPCR assays were performed at Cancer Research UK, Lincolns Inn Fields, by kind invitation and assistance of Dr Charles Swanton and his PhD student, Alvin Lee. All reactions were carried out as per the manufacturers instructions using 50μl TaqMan Universal PCR Master Mix (2X) (Applied Biosystems Inc, Austin, TX, USA) and 37.5ng cDNA made up with water to a total volume of 100μl in each fill reservoir. After application of the mastermix into the fill reservoir, the cards were centrifuged twice for 1 min each at 2200 rpm to evenly distribute the mastermix into each reaction chamber. The card was sealed as per the manufacturers instructions and transferred to a 7900HT Real Time PCR system (Applied Biosystems Inc,

Austin, TX, USA). Thermal cycling conditions were set at 95^0C for 10 mins followed by 40 repeats of 95^0C for 15s and 60^0C for 1min.

45 genes (Tables 16 and 19) were tested in duplicate using the same source RNA used for the microarray analysis (n=9) as well as a second validation set using RNA isolated from a fresh sample set (n=11). The qPCR cards also included 3 reference genes (GAPDH, 18S and ACTB) for normalisation. Raw threshold cycle (Ct) data were extracted using RQ Manager v1.2 software (Applied Biosystems Inc, Austin, TX, USA). Relative quantification of gene expression was calculated using the $\Delta\Delta$Ct method in Microsoft Excel after pooling cancer versus benign samples with expression of genes in the benign set normalised to 1 using 18S or GAPDH as the calibrator.

Table 19. Layout and gene reference data for the customised TaqMan Array format 48. Each column of the table represents the 2 columns of qPCR reaction wells filled by each of the 8 fill reservoirs on the TaqMan Array cards.

POU5F1 Hs00999632_g1	GAPDH Hs00266705_g1
LAMC2 Hs01043711_m1	ACTB Hs00357333_g1
RAB27A Hs00608302_m1	MUC4 Hs00366414_m1
CSPG4 Hs00426981_m1	MUC5AC Hs01365616_m1
PRKCB1 Hs00176998_m1	CD9 Hs00233521_m1
PTK2 Hs00178587_m1	LUM Hs00158940_m1
MYC Hs99999003_m1	MAML2 Hs00418423_m1
LIF Hs00171455_m1	ITGB8 Hs01110394_m1
SERPINA3 Hs00153674_m1	VEGFA Hs99999070_m1
CELSR1 Hs00183906_m1	PNMA2 Hs00246721_s1
ATP6V0A2Hs00429389_m1	18S Hs99999901_s1
NOTCH3 Hs00166432_m1	CEACAM1 Hs00236077_m1
COL1A1 Hs00164004_m1	MXRA5 Hs00377849_m1
COL17A1 Hs00990073_m1	HOXA10 Hs00172012_m1
MCM4 Hs00381539_m1	HOXB6 Hs00255831_s1
MAPK1 Hs01046830_m1	ITIH5 Hs00228960_m1
RAD51 Hs00153418_m1	ASPHD1 Hs00736180_m1
NCSTN Hs00299716_m1	TGFB1 Hs00932734_m1
STAT1 Hs01014005_m1	TM4SF18 Hs00298933_m1
TRIB2 Hs00222224_m1	PVT1 Hs00413039_m1
KRAS Hs00364282_m1	LEF1 Hs00212390_m1
MAML3 Hs00298519_s1	CFDP1 Hs01041483_m1
PEAR1 Hs01378394_m1	MMP2 Hs01548727_m1
COL6A3 Hs00915120_m1	SPARC Hs00234160_m1

6.3 Results

6.3.1 Testing of mRNA isolated from biliary brushings.

This was done as outlined in Chapter 4. In summary, RNA was quantified using the NanoDrop ND-1000 Spectrophotometer and quality assessed by Agilent Bioanalyzer. Prior to use on the TaqMan Array cards, samples were assessed for contamination with genomic DNA and functional analysis of the RNA was performed using single gene SYBR Green qPCR alongside reverse transcriptase negative (RT-) controls using highly expressed genes such as GAPDH or 18S. Samples were considered suitable if there was appropriate amplification of the reference genes and CT values >35 in the RT- controls.

6.3.2 Failed TaqMan Array cards and exclusion of samples with poor amplification thresholds

In the batch of 10 TaqMan Array cards, 5 failed to provide data due to machine misalignment. The majority of these samples were repeated using spare RNA converted to cDNA. Two samples (1 BTC, and 1 benign) were excluded from the second fresh validation set in view of reference gene CT values being much higher (i.e. lower cDNA input) than the other samples resulting in most of the other genes of interest shifting to the right with lack of amplification within the 40 cycles. These samples could therefore not be used for meaningful results and were excluded from further data analysis.

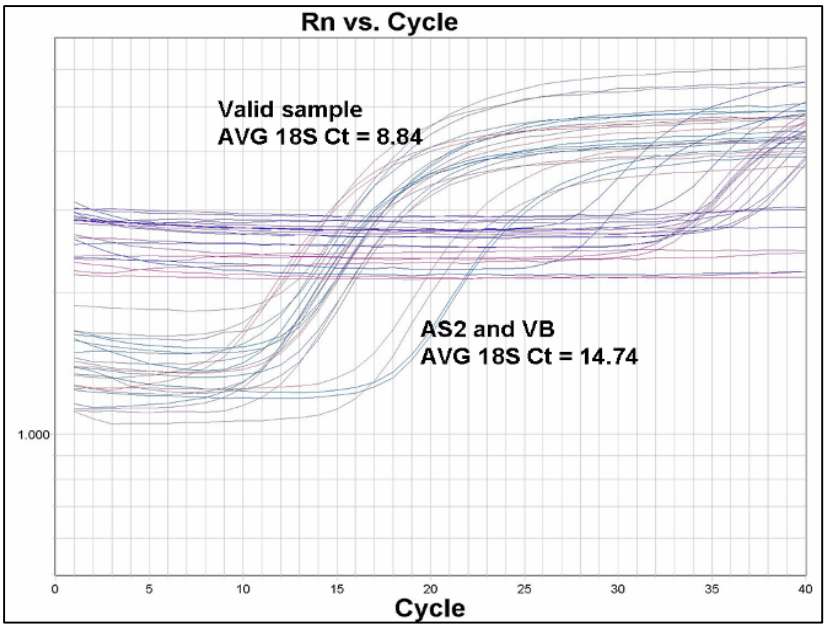


Figure 45. Example of gene expression from 2 samples not demonstrating amplification within the 40 cycles with a right shift in the amplification curves due to lower input levels of cDNA. These 2 samples were excluded from further data analysis.

6.3.3 Relative gene expression assessed by TaqMan Array qPCR

Mean CT and mean ΔCt were used to calculate the relative gene expression as described in Chapter 4. Examples of amplification plots and ΔCt values for a selection of the genes are shown in figure 46.

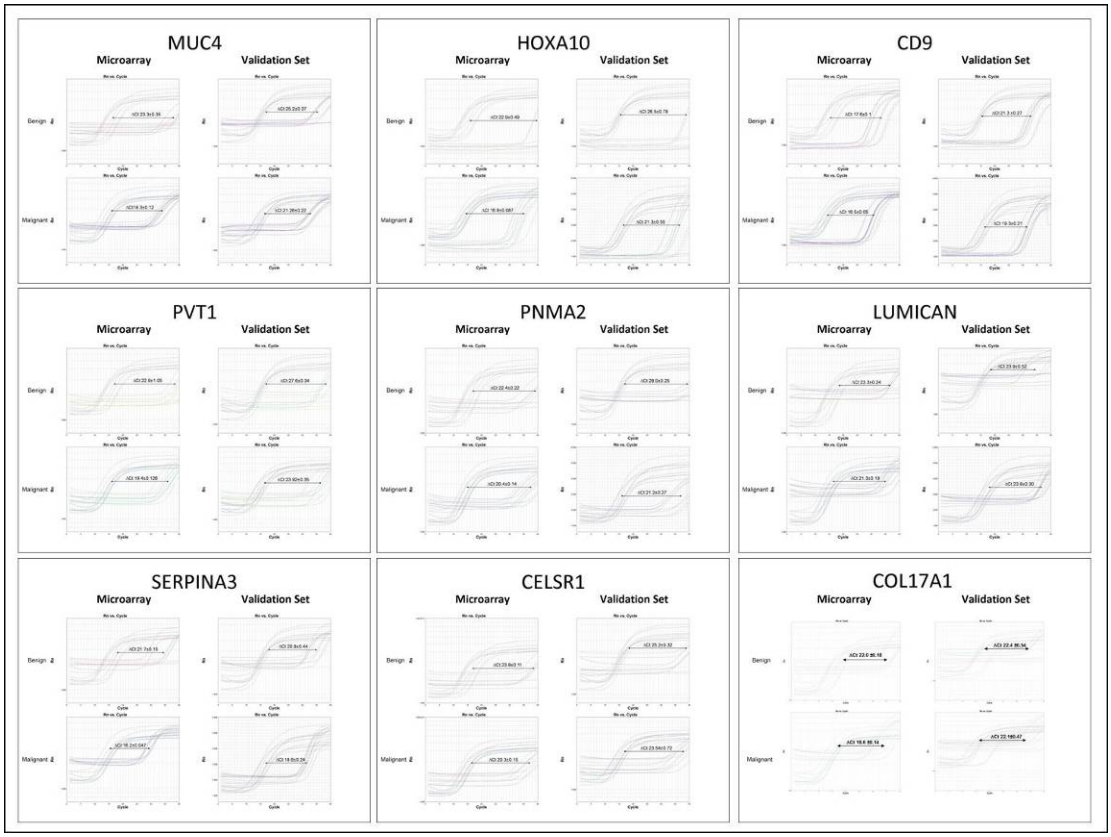


Figure 46. Examples of qPCR amplification plots for 9 of the 45 genes assessed by TaqMan Array. Plots show grouped benign and cancer data normalised to 18S RNA for the microarray validation set 1 and the fresh validation sample set with relative gene expression shown by the mean ΔCt calculations.

Overall, the pattern of gene expression supported that identified by microarray. Exact duplication of levels of fold change were unlikely and not required for confirmation of the trend in gene expression. Concordance of expression pattern was considered present if the fold change was ≥2 in both the microarray and qPCR data.

6.3.4 Validation set 1 (same source RNA used for the microarray experiments)

Up regulated genes analysed using the customised TaqMan Array qPCR cards are shown in the results Table 20 and Figure 47. The pattern of gene

expression was similar in both the microarray and qPCR data with 79% (34/43) concordance of overall gene expression between the 2 platforms when comparing up-regulated genes. Raw Ct data and ΔΔCt calculations of relative gene expression are shown in the Appendix.

6.3.5 Validation set 2 (fresh RNA sample set)

Using the second, fresh validation set, 36 out of 43 up regulated genes showed a similar pattern in gene expression (83% concordance). Levels of fold change are shown in Table 20. Raw Ct data and ΔΔCt calculations of relative gene expression are shown in the Appendix.

Table 20. Results of gene expression assessed by microarray and TaqMan Array qPCR showing mean fold change between malignant versus benign biliary brushings

Gene symbol	Gene ID	Fold change		
		Microarray analysis	qPCR validation set 1	Second, fresh qPCR validation set
ASPDH1	ASPHD1-Hs00736180_m1	**10**	**1.7**	**3.6**
ATP6VOA2	ATP6V0A2-Hs00429389_m1	**10**	**1.1**	**3.1**
CD9	CD9-Hs00233521_m1	**19**	**2.2**	**4.0**
CEACAM1	CEACAM1-Hs00236077_m1	**3.5**	**2.1**	**3.1**
CELSR1	CELSR1-Hs00183906_m1	**5.2**	**7.2**	**7.3**
CFDP1	CFDP1-Hs01041483_m1	**5.3**	**1.1**	**1.7**
COL17A1	COL17A1-Hs00990073_m1	**16**	**10.9**	**4.9**
COL1A1	COL1A1-Hs00164004_m1	**7.2**	**21.6**	**3.4**
COL6A3	COL6A3-Hs00915120_m1	**8.8**	**26.7**	**2.3**
CSPG4	CSPG4-Hs00426981_m1	**8.5**	**19.5**	**1.6**
HOXA10	HOXA10-Hs00172012_m1	**24**	**67.0**	**35.8**
HOXB6	HOXB6-Hs00255831_s1	**20**	**5.0**	**10.2**
ITGB8	ITGB8-Hs01110394_m1	**2.7**	**4.0**	**2.6**
ITIH5	ITIH5-Hs00228960_m1	**20**	**12.4**	**1.5**
LAMC2	LAMC2-Hs01043711_m1	**2.5**	**2.8**	**3.3**
LEF1	LEF1-Hs00212390_m1	**15**	**7.1**	**4.9**
LIF	LIF-Hs00171455_m1	**4.7**	**7.5**	**2.0**
LUM	LUM-Hs00158940_m1	**11**	**3.7**	**1.1**
MAML2	MAML2-Hs00418423_m1	**3.5**	**3.9**	**1.4**
MAPK1	MAPK1-Hs01046830_m1	**2.35**	**1.6**	**2.2**
MCM4	MCM4-Hs00381539_m1	**2.2**	**1.5**	**2.5**
MMP2	MMP2-Hs01548727_m1	**2.6**	**11.0**	**2.3**
MUC4	MUC4-Hs00366414_m1	**3.2**	**29.4**	**15.5**
MUC5AC	MUC5AC-Hs01365616_m1	**3.6**	**7.4**	**5.0**
MXRA5	MXRA5-Hs00377849_m1	**24**	**10.4**	**3.5**
MYC	MYC-Hs99999003_m1	**2.7**	**6.0**	**11.5**
NCSTN	NCSTN-Hs00299716_m1	**3.4**	**1.0**	**1.9**
NOTCH3	NOTCH3-Hs00166432_m1	**6**	**8.0**	**4.7**
PEAR1	PEAR1-Hs01378394_m1	**10.3**	**6.6**	**2.7**
PNMA2	PNMA2-Hs00246721_s1	**39**	**3.9**	**112**
POU5F1	POU5F1-Hs00999632_g1	**3.3**	**2.5**	**5.5**
PRKCB1	PRKCB1-Hs00176998_m1	**8.3**	**6.8**	**3.1**
PTK2	PTK2-Hs00178587_m1	**3**	**1.9**	**1.8**
PVT1	PVT1-Hs00413039_m1	**11**	**12.1**	**12.5**
RAB27A	RAB27A-Hs00608302_m1	**2.5**	**2.7**	**2.7**
RAD51	RAD51-Hs00153418_m1	**4.3**	**3.0**	**4.2**
SERPINA3	SERPINA3-Hs00153674_m1	**37.8**	**46.7**	**4.4**
SPARC	SPARC-Hs00234160_m1	**7.1**	**41.4**	**3.1**
STAT1	STAT1-Hs01014005_m1	**2.5**	**1.9**	**2.8**
TGFBI	TGFBI-Hs00932734_m1	**12.3**	**3.3**	**2.5**
TM4SF18	TM4SF18-Hs00298933_m1	**10.9**	**0.9**	**5.3**
TRIB2	TRIB2-Hs00222224_m1	**4.3**	**3.4**	**3.1**
VEGFA	VEGFA-Hs99999070_m1	**2.5**	**2.6**	**2.9**

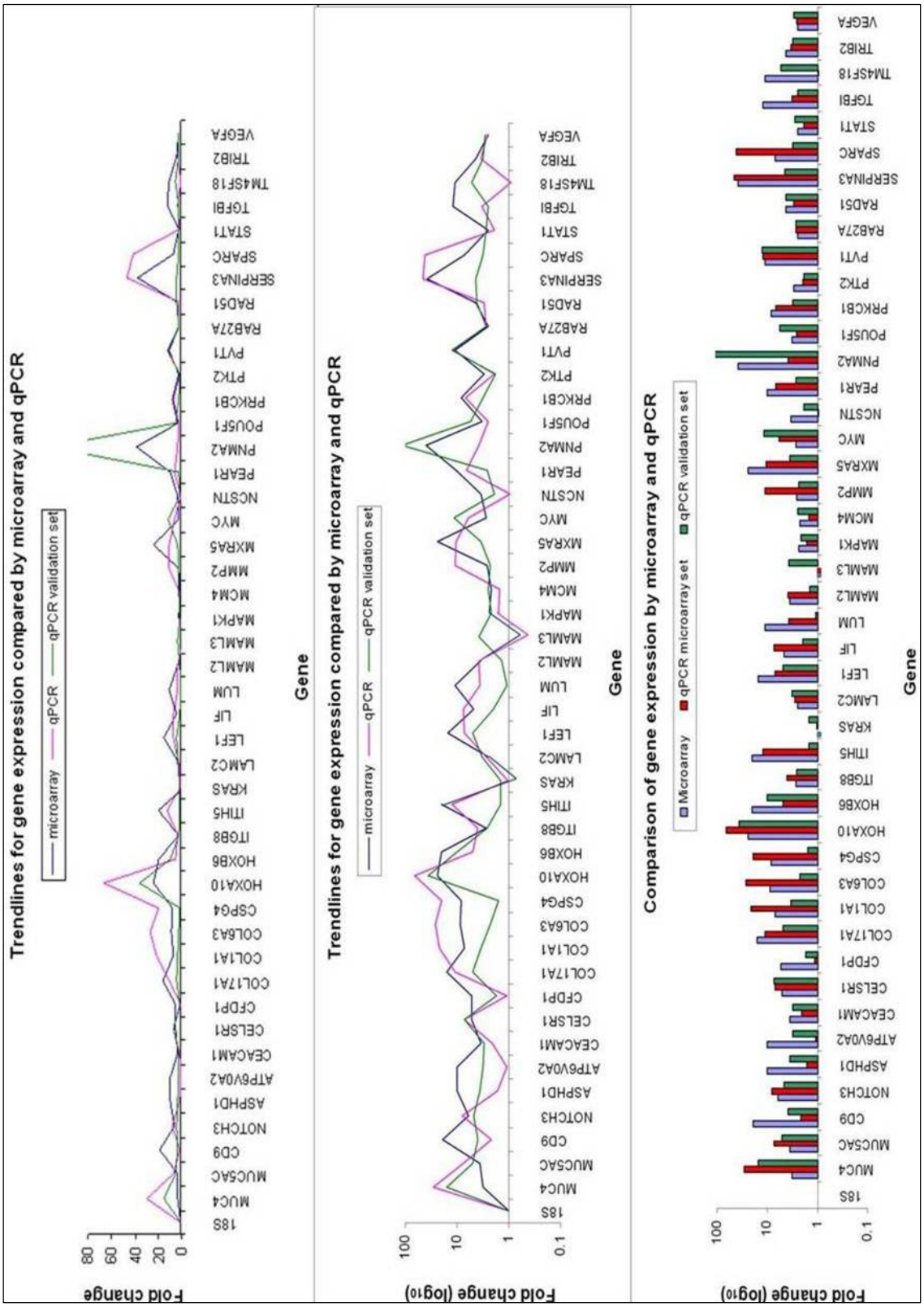


Figure 47. Summary of gene expression data plotted as comparisons between microarray, qPCR validation and fresh qPCR validation sets. The 3 graphs show the same data in different formats to demonstrate the overall trends between different sample sets

6.4 Conclusions: Validation of gene expression using TaqMan Array qPCR

Overall, the upregulation of gene expression identified by microarray was confirmed using qPCR as an alternative method of measuring mRNA expression. These data support further investigation at the protein level for potential biomarkers in BTC. However, multiple gene qPCR by TaqMan Array is still a relatively new and little used assay and therefore we chose to further assess expression of selected genes by standard single gene qPCR.

6.5 Part 2: Single gene qPCR validation of mRNA expression.

6.6 Introduction

In order to further validate the gene expression pattern using an alternative technique, we assessed the mRNA expression of a small number of genes by the SYBR Green qPCR technique. The genes assessed were:

1. CD45 and CK19 markers were used to assess the relative quantities of leukocyte and epithelial cell RNA in biliary brush samples.
2. CD9 and LUM – their expression was significantly upregulated in the microarray data with no overlap in benign and malignant samples and published data on suitable, tested primer sets were available.
3. MUC4 and MUC5AC – reliable and previously tested primer sets were available in our laboratory. MUC4 was used as an internal control of up-regulation in cancer as previously demonstrated (Matull, Andreola et al. 2008)

6.6.1 CD9

Cluster of Differentiation 9 (CD9, also known as MRP-1, MIC3, TSPAN29, P24, 5H9, BA2, BTCC-1, DRAP27, and GIG2) is a member of the transmembrane 4 super family. It is expressed by epithelial cells, endothelial cells, neurons and vascular smooth muscle cells. CD9 expression is high in bone marrow derived progenitor cells but low in activated lymphocytes and myeloid cells and is not expressed in resting T or B lymphocytes (NCBI database information). Data from the Human Protein Atlas (www.proteinatlas.org) show that CD9 is not present in normal cholangiocytes on immunohistochemical staining. CD9 contributes to normal cell adhesion and motility,

and up-regulation may play a role in invasiveness and metastasis of some cancers including melanoma, prostate and cervical cancer, although its expression is reported to be inversely correlated with poor prognosis in lung, pancreas and breast cancers (Longo N 2001, Sauer G 2003, Miyake M, 1995). Mixed reports on CD9 expression in cancer may relate to the observation that cells in contact with the basement membrane, advancing tumour edges and areas of lymphovascular invasion have elevated expression of CD9 (Sauer G 2003, Longo N 2001, whereas it may be reduced in the centre of tumour masses. Elevated expression in BTC may relate to the tendency of cholangiocarcinoma to infiltrate along bile ducts rather than form mass lesions.

6.6.2 Lumican

Lumican (LUM, also known as LDC or SLRR2D) is a member of the small leucine rich proteoglycan (SLRP) family. The protein is expressed in the extracellular matrix of the pancreas, liver, skin, cornea, heart, placenta and skeletal muscle but not in normal biliary epithelium or lymphoid tissue (data from NCBI databases and the Human Protein Atlas). Lumican is involved in modulation of collagen synthesis, tissue repair and control of cellular proliferation and migration (Chakravarti S 2002). Studies in some cancers including colonic and pancreatic cancer demonstrate elevated levels of lumican in malignant disease (Lu YP 2002, Koninger J 2004). Our microarray data suggest a significant upregulation of lumican in BTC (fold change 11) and in view of its location as an extracellular matrix protein, it is plausible that lumican fragments may be identified in the circulation as a tumour marker and deserve further investigation in BTC.

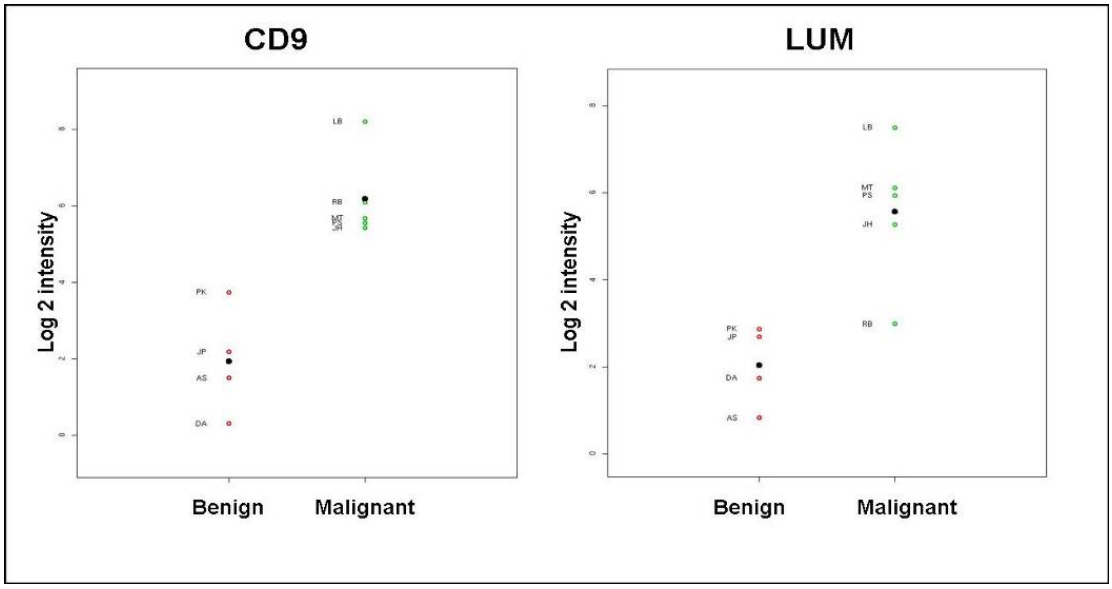


Figure 48. Microarray gene expression plots showing significant elevation of CD9 (mean fold change 19) and lumican (fold change 11) in BTC compared to benign biliary disease.

6.7 Materials and methods

6.7.1 Samples

Clinical samples of biliary brushings were those described in Chapter 4. The same purified RNA samples were used as that used for the TaqMan Array cards in order to provide a direct comparison of results. A summary of the 11 samples used is shown in Table 21.

Table 21. Samples used for further qPCR validation using the SYBR Green method.

	Sample	Diagnosis	Age	Gender
Benign	**B5**	PSC	35	M
	B6	CBD Stricture and Stones	65	M
	B7	IgG4 Disease	69	M
	B8	SOD	38	F
	B9	IgG4 Disease	47	M
			Mean 53	
Malignant	**M6**	CC	79	M
	M7	CC	63	M
	M8	CC	60	F
	M4	CC	59	M
	M9	CC	71	F
	M10	CC	72	M
			Mean 67	

6.7.2 RNA isolation, purification and cDNA synthesis.

RNA from the biliary brushings was isolated and purified as described in chapter 4. cDNA was synthesized using the High Capacity RNA to cDNA Synthesis Kit (Applied Biosystems Inc, Austin, TX, USA) as per the manufacturers instructions. cDNA synthesis was performed in 20μl volumes from a template of 150ng total RNA using the reverse transcriptase enzyme mix for 60mins at 37^{o}C. Reverse transcriptase negative (RT-) controls were made for the majority of samples where there was potential spare RNA in

order to test for genomic DNA contamination. cDNA samples were stored at -20^{o}C until further use.

6.7.3 SYBR Green qPCR

qPCR reactions were carried out in 25μl reaction volumes in 96 well plates on an Applied Biosystems 7500 Real Time PCR machine (Applied Biosystems Inc, Austin, TX, USA). Experiments were performed in duplicate ordinarily using cDNA input equivalent to 7.5ng RNA (range for some experiments 5ng-10ng equivalent RNA converted to cDNA) with the addition of SYBR Green qPCR Mastermix (x2) with low Rox (Eurogentec) and 0.5μl each of 10μM forward and reverse primer pairs.

Primer pair sequences used for the SYBR Green qPCR reactions are shown in Table 22. Primer sets were chosen based on short amplicon lengths (as discussed in Chapter 4) as well as previous demonstration of suitability for SYBR Green qPCR. Primer sets for CD9 and LUM were synthesized and purchased from Eurogentec as per the sequences shown in the table. Efficiency of amplification was assessed using a standard curve of increasing concentrations of RNA isolated from human TFK-1 BTC cells in culture (Figure 50).

Table 22. qPCR primer sets used for SYBR Green qPCR analysis of gene expression in biliary brush and bile samples.

Primer	Amplicon length (bp)	Forward sequence (5' - 3')	Reverse sequence (5' - 3')	Reference
GAPDH 87 NM_002046	87	556TGCACCACCAACTGCTTAGC575	642GGCATGGACTGTGGTCATGAG621	(Wu, Lin et al. 2007)
18S RNA NR_003286.2	99	331CGGCGACGACCCATTCGAAC350	429GAATCGAACCCTGATTCCCCGTC407	BIORAD sequence
MUC4 AF058803	101	1499GCCCAAGCTACAGTGTGACTCA1520	1600ATGGTGCCGTTGTAATTTGTTGT1578	(Matull, Andreola et al. 2008)
CD9 NM_001769.2	78	498CAACAAGCTGAAAACCAAGGA 518	576CAAACCACAGCAGTTCAACG 557	(Gandemer, Rio et al. 2007)
LUM NM_002345.3	97	489TCACCAAACTGTGCACCAGAA509	GGAGGCACCATTGGTACACTTT565	(Boheler, Volkova et al. 2003)

6.8 Results

6.8.1 Testing of qPCR primer sets

Prior to use with clinical samples, primer sets require testing to ensure appropriate and specific amplification of PCR products. The primer sets for GAPDH, 18S RNA and MUC4 had previously been tested within the group (Dr Fausto Andreola) and shown to be suitable for qPCR using the SYBR Green method. Examples of the single dissociation peaks are shown in Figure 49.

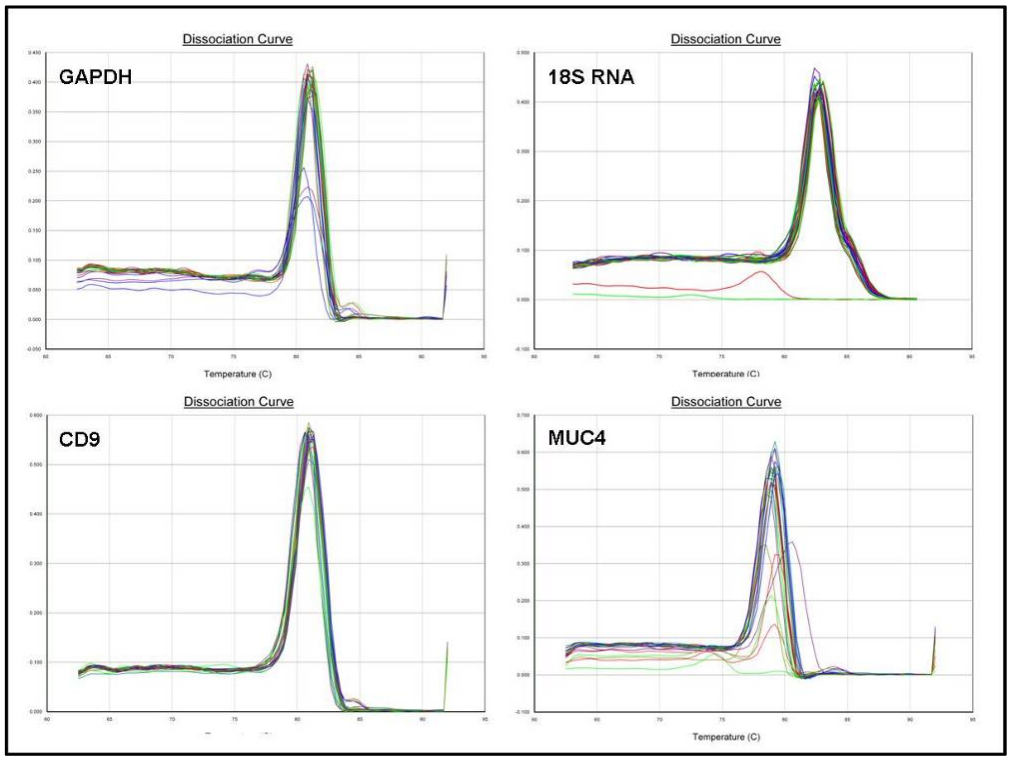


Figure 49. qPCR product dissociation plots showing single, specific peaks in keeping with absence of non-specific gDNA amplification or primer dimerisation.

Efficiency of primer sets was tested using a standard curve of increasing amounts of input control (TFK-1 cell) RNA against the amplification threshold (ΔCt). Results should show a direct correlation. As shown in Figure 50, both the CD9 and LUM primer sets performed satisfactorily.

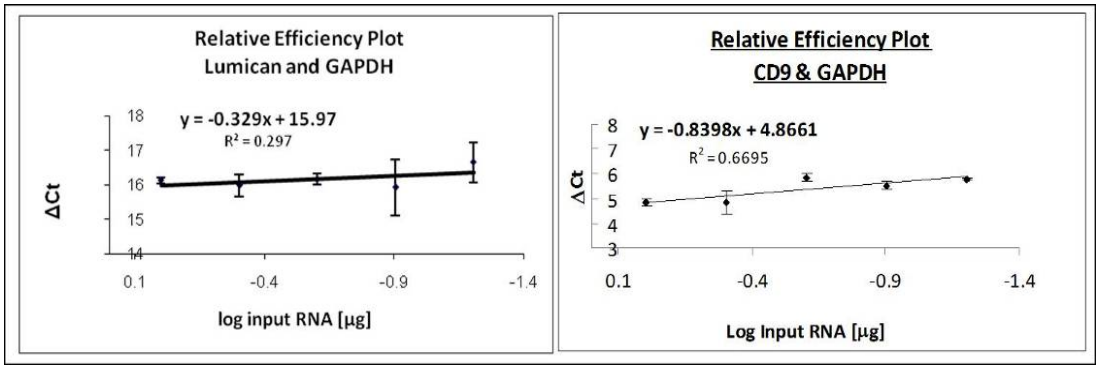


Figure 50. Standard curves for CD9 and LUM primer sets to assess efficiency of PCR amplification. Both primer sets produced satisfactory standard curves.

Specificity for mRNA amplification is tested most simply by evaluating the dissociation plot after completion of the qPCR cycles. The CD9 dissociation plot showed a single specific dissociation peak but the LUM dissociation plots demonstrated non specific additional peak suggestive of primer dimerisation that would affect the accuracy of the gene expression data.

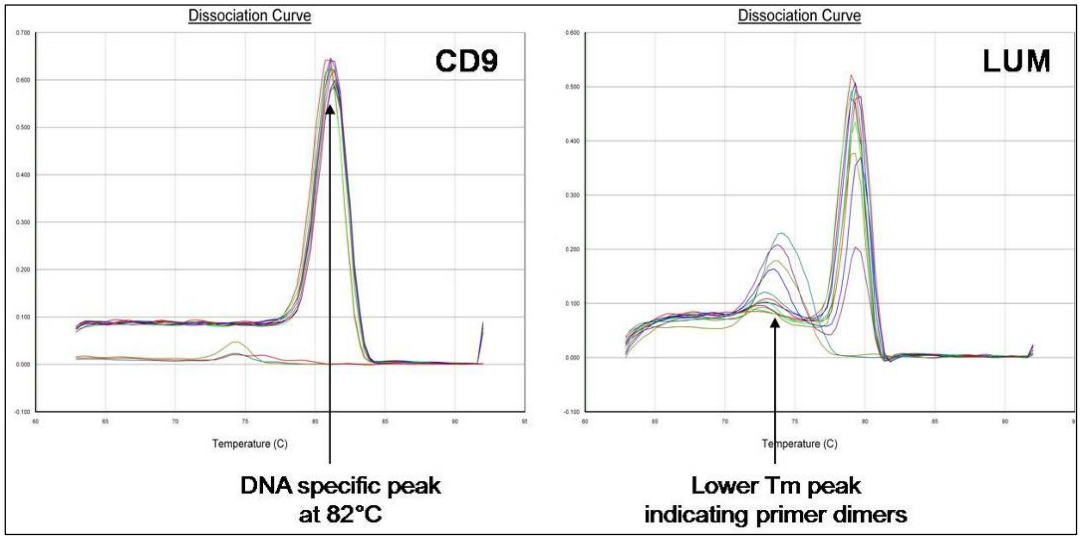


Figure 51. SYBR Green qPCR dissociation plots showing a single specific peak for CD9 but a second, non-specific peak of primer dimerisation using the LUM primer set. (image from Rob Tidswell).

Further experiments were performed using varying concentrations of forward and reverse primers as well as $MgCl_2$ which inhibits non-specific hydrogen bonds and may prevent primer dimerisation. None of the conditions removed the non specific peaks. These studies were performed by a supervised iBSc student (Robert Tidswell). No further analysis of LUM expression by SYBR Green qPCR was undertaken (data not shown).

6.8.2 Relative expression of CK19 and CD45 as markers of epithelial and leukocyte RNA respectively.

These data have already been presented in Chapter 4. In summary, there was no significant difference in relative expression of CK19 or CD45 between benign and malignant samples, although there was a trend towards higher CD45 expression in malignant samples. Also, CK19 mRNA expression was much higher than CD45 mRNA demonstrating that epithelial cells represent the major source of RNA in the biliary brush samples (Figure 52).

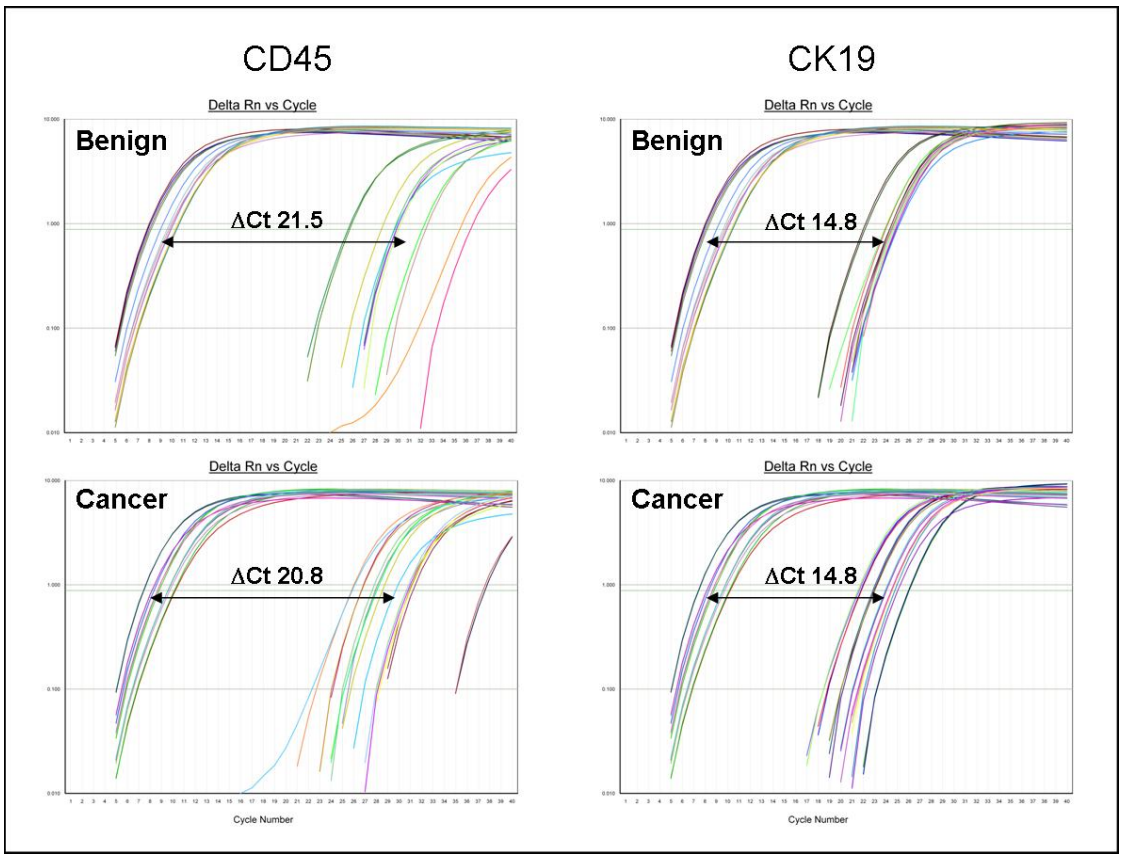


Figure 52. qPCR amplification plots for the pan leukocyte marker CD45 and biliary epithelial cell marker CK19 relative to the reference gene 18S, showing a much greater proportion of epithelial than leukocyte RNA in biliary brushings. Also, there is no significant difference in relative expression of each of the markers between benign and malignant samples. There are much wider variations in leukocyte compared to epithelial RNA quantities as shown by the wide spread in amplification curves and Ct values for CD45 between different samples.

6.8.3 Expression of Lumican in biliary brushings

The lumican primer sets were considered unsuitable for reliable qPCR so were not investigated further in clinical samples in significant numbers.

6.8.4 Expression of CD9 in biliary brushings and bile

When assessed by SYBR Green qPCR, CD9 was shown to be upregulated in cancer versus benign samples by 3 different experiments using biliary brush RNA (n=6 to n=11 per experiment). In the largest experiment (n=11), CD9 had a fold change of 2.8 [95% CI 2.6 to 3.0] in cancer versus benign. This level of upregulation in smaller than that reported by microarray (fold change of 19) but still demonstrates confirmation of upregulation at the mRNA level in the malignant biliary brushings. Examples of individual sample qPCR plots are shown in Figure 53.

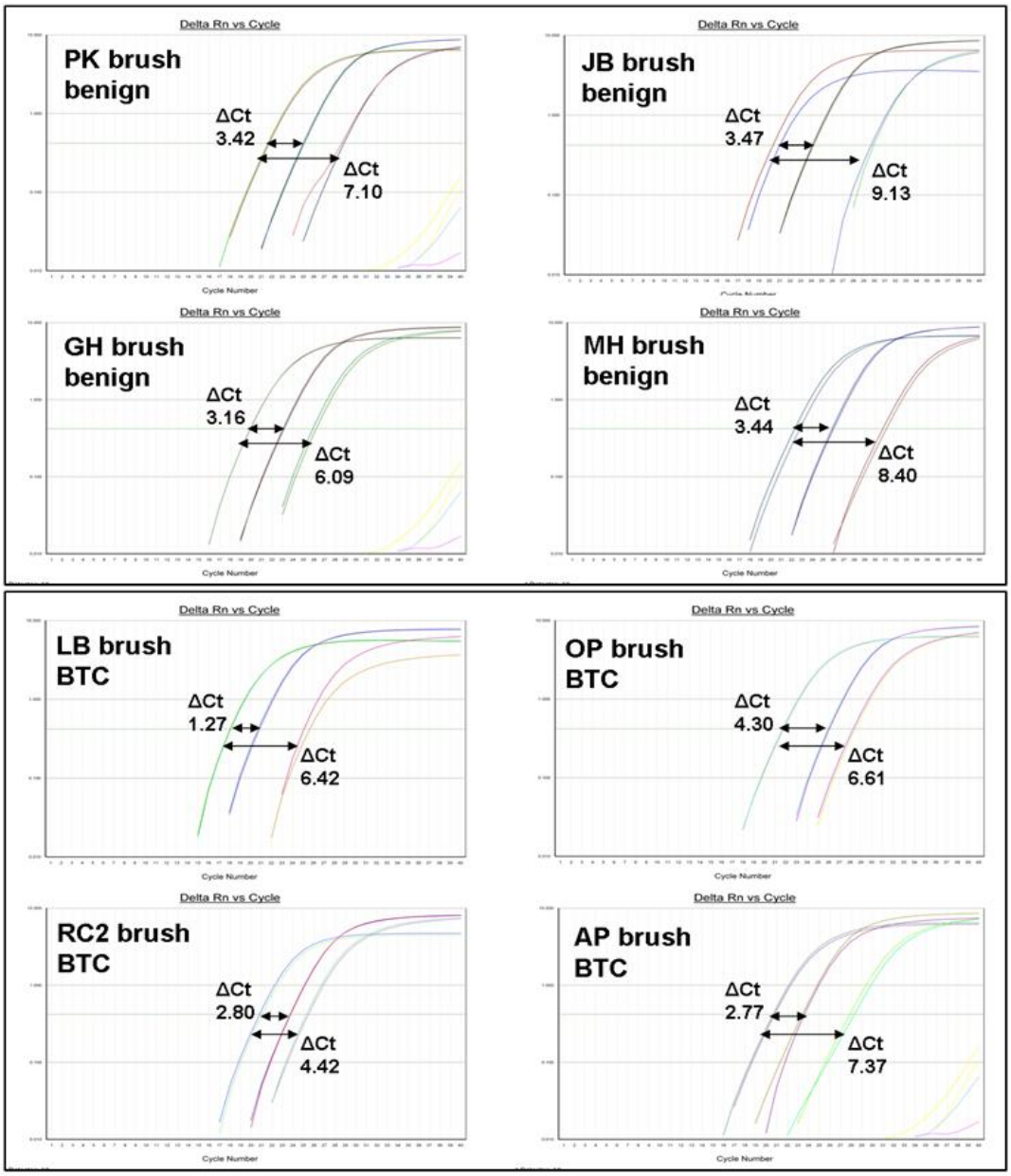


Figure 53. Representative qPCR amplification plots and Ct values for CD9 (middle plot on each graph) and MUC4 (right hand plot on each graph) using GAPDH as the reference gene. The ΔCt are on average smaller (i.e. higher expression) in cancer versus benign samples for both CD9 and MUC4.

A small number of bile samples were used to assess the expression of CD9 relative to GAPDH. Numbers were too small (n=6) for definitive results but suggest an upregulation of CD9 in bile from patients with BTC (ΔCt 4.14 in

cancer versus 6.08 in benign disease); the fold change was 3.8 in cancer versus benign disease. Examples of CD9 qPCR plots are shown in Figure 54.

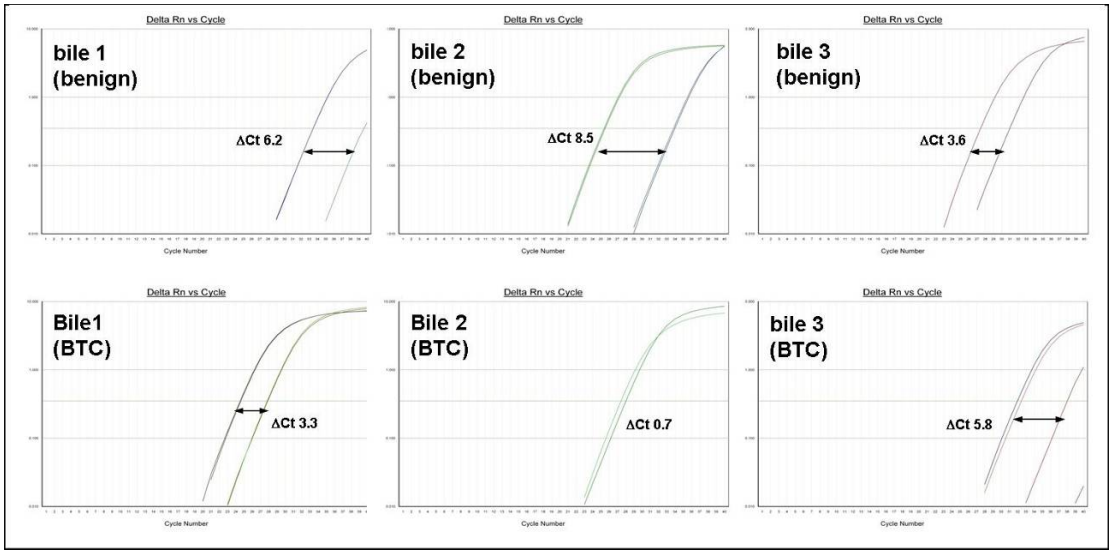


Figure 54. Examples of qPCR plots for CD9 expression in bile from patients with benign and malignant biliary diseases. The mean ΔCt value is smaller in malignant samples suggesting higher expression of CD9 in BTC.

6.8.5 Expression of MUC4 and MUC5AC in biliary brushings

MUC4 was tested as an internal control as it has already been shown to be upregulated at mRNA level in bile and protein level in tissue (Matull, Andreola et al. 2008). As shown in Figure 55, there was a wide variation in MUC4 expression levels in both benign and malignant samples, a phenomenon that was also noted in the microarray results.

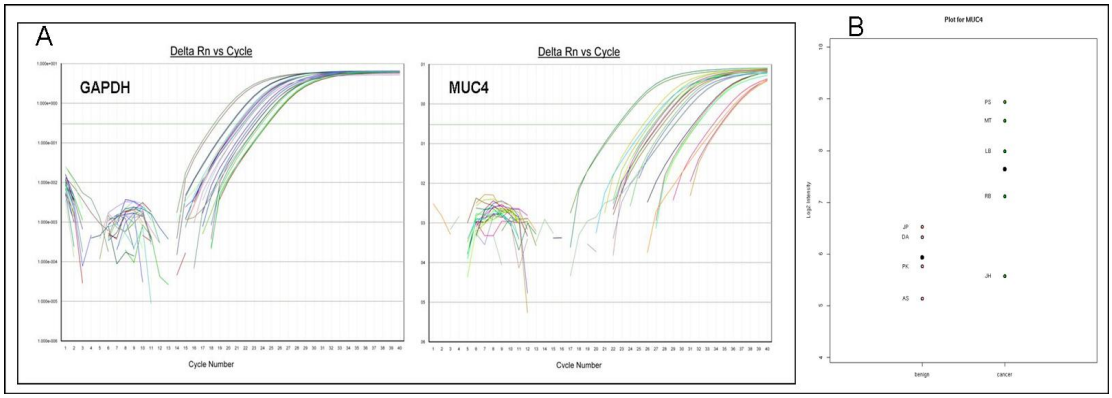


Figure 55. Plots demonstrating a high variability in MUC4 mRNA expression in biliary brush samples by A) qPCR and B) microarray.

Validation using a further set of biliary brush samples by individual gene qPCR using the SYBR Green detection method confirmed upregulation of MUC4 (fold change 21.4 [95% CI 19.1 to 24.1], n=9). Up-regulation of MUC4 is shown above in Figure 53 and acts as a further internal control providing further validity to the up-regulation of CD9 shown at the same time in the same sample set.

MUC5AC was tested in only a very small sample set (n=4) insufficient for proper data analysis. However, the qPCR amplification plots and Ct values show that MUC5AC is highly expressed with Ct values only slightly higher than GAPDH (Figure 56). The range of relative expression was 0.2 to 1.6 but numbers were insufficient to calculate differences in benign and malignant samples or make further conclusions.

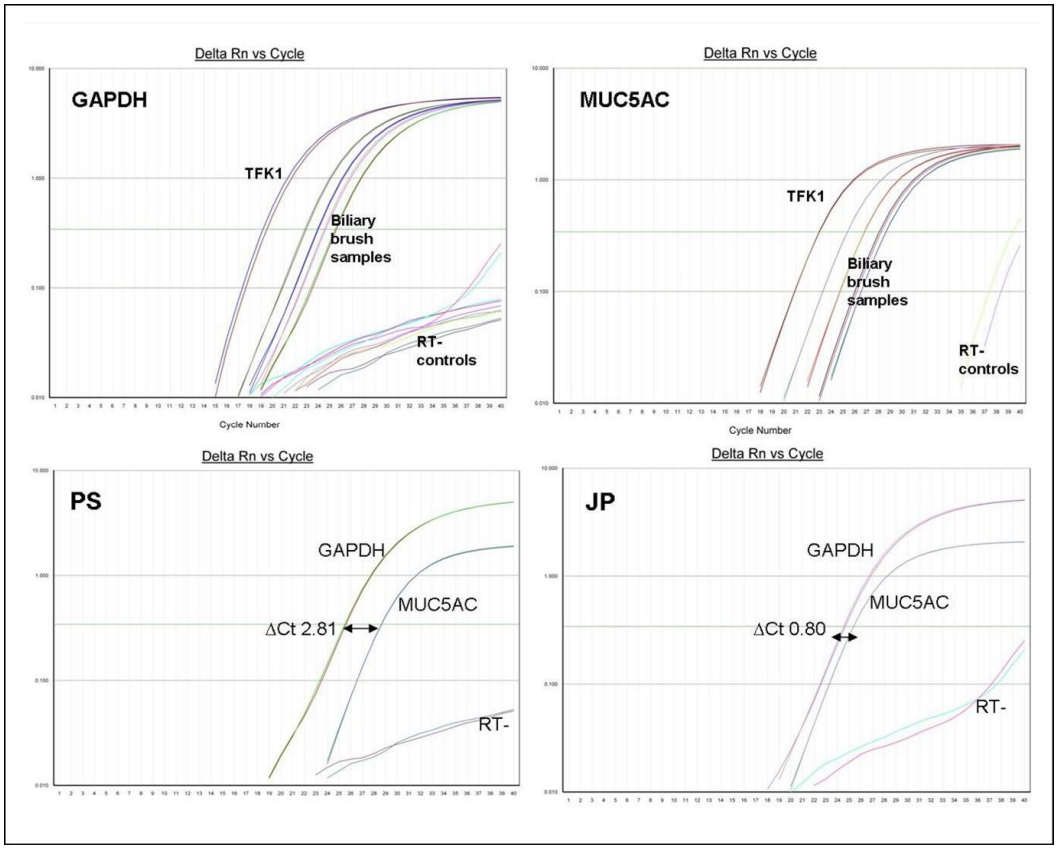


Figure 56. Expression of MUC5AC in biliary brush samples. Note expression levels near to that of GAPDH (DCt 0.8 to 2.8) suggesting high expression in benign (bottom right) and malignant (bottom left) biliary brushings.

6.9 Conclusions

The main conclusions drawn from data presented in this chapter are:

- Where measured, gene expression in biliary brushings assessed by qPCR, has a high concordance (79-83%) with that demonstrated by microarray.
- Elevation of MUC4, previously shown to be upregulated in BTC, was again identified as upregulated in our samples of biliary brushings, acting as a type of internal positive control to support our results.
- These data support the validitity of the microarray data as representative of gene expression in cells isolated from biliary brushings
- The primary source of RNA in biliary brushings is from epithelial cells with much more minor quantities from leukocytes. Therefore, confounding from bile duct infections may be present but would be small.
- These data provide support for further investigation of biliary strictures by microarray and qPCR in greater numbers and also to investigate other biliary diseases such as PSC and IgG4 disease.
- Elevated mRNA expression in genes of interest deserves further investigation to assess the expression of the protein products as potential biomarkers in BTC.

6.10 Discussion

In this chapter we demonstrate that the majority of genes identified by microarray analysis as upregulated in biliary brushings of BTC are also shown to be up-regulated at the mRNA level by qPCR. The actual level of fold change varies depending on the assay used but one cannot expect the data to be exactly the same in all assays and when using relatively small numbers of samples with large biological variations. If the sample size were increased, then it is highly likely that overall variation would fall. The aims of these studies were to identify novel biomarkers, primarily at the protein level. The level of mRNA expression does not necessarily directly translate to the level of protein expression. Reasons for variation between gene and protein expression include post translational modification by microRNA and other means, or non coding gene mutations in the genes of interest and a wide variation in the half life of protein products. Therefore, precise levels of gene expression are less important than confirming a definite trend in fold change so that one can then study relevant protein expression levels directly. This will be addressed in Chapter 7.

As described, these data support the use of global gene expression profiling by microarray in biliary brushings. Some other biliary diseases such as PSC and IgG4 disease, are difficult to study at the cellular and molecular level because of a paucity of clinical materials and commonly an absence of biopsiable tissues. The methods described so far, may therefore be well suited to study such biliary diseases in more depth.

Chapter 7.

7 Analysis of gene and protein expression in biliary tissues

7.1 Introduction and aims

In Chapters 5 and 6 we identified some of the genes upregulated in biliary brushings taken from patients with BTC. Cells brushed at the surface of strictures at ERCP may include a mixture of cell types but are primarily epithelial cells. These cells may be benign and/or malignant and may contain a number of dead cells or cells undergoing apoptosis at the tumour edge exposed to bile. Also, cells at the surface of tumours may have different cell biology to those at the central zones, a phenomenon that is well recognised in tumour biology. However, fresh frozen surgical tissues are still regarded as the best tissues for further study with gene expression analysis and so similarities between tissue and biliary brushings would provide further supportive evidence for biliary brushings as clinical samples suitable for translational work.

In this chapter we aimed to assess;

1. whether a similar gene expression pattern is present in resected biliary tract cancer tissue from patients undergoing attempted curative surgery
2. whether the upregulated gene expression is translated to upregulated protein expression in biliary tissues in order to assess whether these may become biomarkers in BTC.

This Chapter will be divided into two parts; the first addressing mRNA expression analysis in the surgical tissues using methodology described in Chapters 4 and 6; the second assessing protein expression in bile and surgical tissue sections.

7.2 Part 1. Analysis of relative mRNA expression in resected biliary tract cancer tissues

7.3 Introduction

The biology of cells at the surface of tumours is somewhat different to those in the necrotic and ischaemic central zones. Also, small and large cholangiocytes at different locations in the biliary tree have different gene and protein expression profiles, a phenomenon best described in animal studies (Ueno, Alpini et al. 2003). Samples of superficial endoscopic biliary brushings are not necessarily the same as samples of tissue taken from within the tumour mass. However, fresh frozen tissue samples should contain good quality RNA and represent the best (although not perfect) control for comparison with the biliary brush data.

In this section, we describe analyses of gene expression in surgical resection material carried out in order to compare with those found in the biliary brush samples as shown in Chapters 5 and 6.

7.4 Materials and Methods

In general, experimental methods used were the same as those used for the biliary brush material described in Chapter 6.

7.4.1 Surgical resection tissue

Clinical samples were collected following written informed consent for ethically approved research (NHNN Ethics Committee, reference 06/Q0152/106). Surgical resection material was collected at the time of operation at University College Hospital with the kind assistance of the surgeons Professor Massimo Malago, Mr Charles Imber and Dr Steven Olde Damink. Once the relevant area of liver, bile duct or tumour had been resected, the surgeon dissected out a small area (approx 0.5-2cm^3) of tissue that was then snap frozen in liquid nitrogen and stored at -80^{0}C until further use. Pathological confirmation of diagnosis in the main resected tissue specimen was confirmed for all cases used. Tissues used for the TaqMan array qPCR are shown in Table 23. In addition, a few other samples were collected and assessed (e.g. normal liver from patients undergoing resection of colorectal metastasis and bile ducts from patients undergoing Whipples procedures for pancreatic cancer) but were generally not used for calculations of relative gene expression and these data are not presented in the main data set.

Table 23. Clinical details for surgical resection material used in the qPCR experiments.

Sample	Diagnosis
Benign1	Benign inflammatory / ischaemic CBD stricture
Benign2	Cystic duct at cholecystectomy for stones
Benign3	Benign inflammatory / ischaemic CBD stricture
Benign4	Cystic duct at cholecystectomy for stones
Benign5	Left hepatic duct at surgery for CRC liver metastasis
Benign6	Benign CBD stricture (IgG4 disease)
BTC1	Distal CBD cholangiocarcinoma
BTC2	Hilar cholangiocarcinoma
BTC3	Hilar cholangiocarcinoma
BTC4	Distal CBD cholangiocarcinoma
BTC5	Hilar cholangiocarcinoma
BTC6	Distal CBD cholangiocarcinoma
BTC7	Hilar cholangiocarcinoma
BTC8	Hilar cholangiocarcinoma
BTC9	Hilar cholangiocarcinoma

7.4.2 RNA isolation, purification and cDNA synthesis.

The surgical resection material was fragmented by crushing in a pestle and mortar whilst frozen in liquid nitrogen or mechanically homogenised in TRI reagent. RNA was isolated using TRI Reagent (Ambion) and further purified by spin column purification (RNEasy Mini, Qiagen) as per the manufacturer's instructions and eluted in 20μl of RNase free water. RNA concentrations and quality were assessed by Nanodrop ND-1000 spectrophotometer (Labtech). cDNA was synthesized using the High Capacity RNA to cDNA Synthesis Kit (Applied Biosystems Inc, Austin, TX, USA) as per the manufacturers instructions. cDNA synthesis was performed in 20μl volumes from a template of 300ng total RNA (double the quantity used for the biliary brush experiments) using the reverse transcriptase enzyme mix for 60mins at 37^{o}C. Reverse transcriptase negative (RT-) controls were made in order to test for genomic DNA contamination. cDNA samples were stored at -20^{o}C until further use.

7.4.3 Single gene qPCR using the SYBR Green assay

qPCR reactions were carried out in 25μl reaction volumes in 96 well plates on an Applied Biosystems 7500 Real Time PCR machine (Applied Biosystems Inc, Austin, TX, USA). Experiments were performed in duplicate ordinarily using cDNA input equivalent to 7.5ng RNA (range for some experiments 5ng-10ng equivalent RNA converted to cDNA) with the addition of SYBR Green qPCR Mastermix (x2) with low Rox (Eurogentec) and 0.5μl each of 10μM forward and reverse primer pairs.

Primer pair sequences for the SYBR Green qPCR reactions were the same as those used in the biliary brush experiments (Table 24). The primary genes assessed were;

- 18S (NR_003286.2) and GAPDH (NM_002046) as reference genes
- CK19 (NM_002276.3) and CD45 (NM_080922.1) to assess relative quantities of epithelial and leukocyte RNA in the tissue samples
- CD9 (NM_001769.2) for comparison with biliary brush and tissue TaqMan Array qPCR results.

Table 24. Primer sets used for SYBR Green qPCR experiments in biliary tissues.

Gene & GeneBank accession number	Amplicon length (bp)	Forward sequence (5' - 3')	Reverse sequence (5' - 3')
GAPDH 87 NM_002046	87	556TGCACCACCAACTGCTTAGC575	642GGCATGGACTGTGGTCATGAG621
18S RNA NR_003286.2	99	331CGGCGACGACCCATTCGAAC350	429GAATCGAACCCTGATTCCCCGTC407
CK19 NM_002276.3	79	514TACAGCCACTACTACACGACCATCC538	592GGACAATCCTGGAGTTCTCAATG570
CD45 NM_080922.1	102	1353GGAAGTGCTGCAATGTGTCATT1374	1454CTTGACATGCATACTATTATCTGATGTCA1426
CD9 NM_001769.2	78	498CAACAAGCTGAAAACCAAGGA 518	576CAAACCACAGCAGTTCAACG 557

As described previously in Chapters 4 and 6, the comparative threshold ($\Delta\Delta$Ct) method was used to calculate relative expression of the target gene compared to the reference gene (usually 18S).

7.4.4 Custom designed TaqMan Array qPCR cards

In order to compare relative BTC gene expression in tissue with that of biliary brushings, we used custom synthesized TaqMan Array qPCR cards with a format similar to that used for the biliary brushings experiments in Chapter 6. As before, cards were designed with a format of 48 genes per fill reservoir. The gene IDs were similar to those used in the previous biliary brush experiments with the difference being replacement of 2 genes with 2 new gene primer-probe sets not used in the first set. A list of genes included in the TaqMan Array cards is shown in Table 21.

The B-actin control was removed in view of the fact that results were not reproducible in the first set, a phenomenon that may be related to the primer-probe set used or variation in cancer versus benign tissue. MAML3 was also removed in view of the fact that the change in gene expression identified by microarray was not found in either of the 2 qPCR validation sets, suggesting a possible false positive result from the microarray data. These two genes were replaced by two other genes on the original shortlist, VIM and MLL4. Vimentin (VIM) is an intermediate filament protein of the cytoskeleton primarily expressed in mesenchymal tissues and over-expression has been reported in a number of other solid organ cancers (NCBI database information). Our

microarray data reported a fold change of 3.4 (p=0.006) in malignant biliary brushings. Despite its name, mixed lineage leukaemia 4 (MLL4) is ubiquitously expressed by all tissues. The protein product is a histone methyltransferase involved in epigenetic transcriptional activation and has been shown to be up-regulated in other solid organ cancers (NCBI database information). Our microarray data reported a fold change of 7.7 (p=<0.001) in cancer versus benign biliary brushings. These 2 genes were added in place of ACTB and MAML3.

Table 25. Genes included in the custom synthesized TaqMan Array qPCR cards used for analysis of gene expression in biliary surgical resection material.

List of genes for qPCR validation using TaqMan arrays					
18S RNA	ITGB8	CELSR1	CFDP1	ATP6V0A2	STAT1
GAPDH	VEGFA	NCSTN	MMP-2	PRKCB1	NOTCH3
MLL4	PNMA2	ASPHD1	COL6A3	PTK2	COL1A1
MUC4	CEACAM1	TGFBI/1	SPARC	cMYC	BPAG2
MUC5AC	MXRA5	TM4SF18	TRIB2	SERPINA3	SEL1L
CD9	HOXA10	PVT1	LAMC2	LIF	MCM4
LUM	HOXB6	LEF1	RAB27A	RAD51	MAPK1
MAML2	ITIH5	POU5F1	CSPG4	VIM	KRAS

TaqMan® Array qPCR assays were performed at Cancer Research UK, Lincolns Inn Fields, by kind invitation and assistance of Dr Charles Swanton and his PhD student, Alvin Lee. All reactions were carried out as per the manufacturers instructions using 50μl TaqMan Universal PCR Master Mix (2X) (Applied Biosystems Inc, Austin, TX, USA) and 75ng cDNA (double the quantity used for the biliary brush experiments) made up to a total volume of 100μl with water in each fill reservoir. After application of the cDNA mastermix into the fill reservoir, the cards were centrifuged twice for 1 min each at 2200 rpm to evenly distribute the mastermix into each reaction chamber. The card

was sealed as per the manufacturers instructions and transferred to a 7900HT Real Time PCR system (Applied Biosystems Inc, Austin, TX, USA). Thermal cycling conditions were set at 95^0C for 10 mins followed by 40 repeats of 95^0C for 15s and 60^0C for 1min.

45 genes (Table 25) were tested in duplicate using RNA isolated from the surgical resection material (n=15). The qPCR cards also included 2 reference genes (18S and GAPDH) for normalisation. Raw threshold cycle (Ct) data was extracted using RQ Manager v1.2 software (Applied Biosystems Inc, Austin, TX, USA). Relative quantification of gene expression was calculated using the $\Delta\Delta$Ct method in Microsoft Excel after pooling cancer versus benign samples with expression of genes in the benign set normalised to 1 using 18S or GAPDH as the calibrator.

7.5 Results

7.5.1 RNA concentrations and quality.

As expected, RNA from fresh frozen resection specimens was of much higher quality and quantity than that isolated from biliary brushings. Examples of spectrophotometric analyses of purified RNA are shown in Figure 57.

Sample ID	User ID	Date	Time	ng/ul	A260	A280	260/280
MJ tumour RNA 11.8.09	Default	11/08/2009	12:58	276.55	6.914	3.959	1.75
MT tumour RNA 11.8.09	Default	11/08/2009	12:59	130.23	3.256	1.595	2.04
DG tumour RNA 11.8.09	Default	11/08/2009	13:00	51.21	1.280	0.592	2.16
DG bile duct RNA 11.8.09	Default	11/08/2009	13:01	5.59	0.140	0.049	2.84
RT tumour RNA 11.8.09	Default	11/08/2009	13:03	38.49	0.962	0.469	2.05
RT bile duct RNA 11.8.09	Default	11/08/2009	13:04	163.11	4.078	1.943	2.10
MF bile duct RNA 11.8.09	Default	11/08/2009	13:05	-2.52	-0.063	0.007	-9.48
SO tumour RNA 11.8.09	Default	11/08/2009	13:06	494.71	12.368	5.704	2.17
LR bile duct RNA 11.8.09	Default	11/08/2009	13:08	260.71	6.518	3.119	2.09
LP bile duct RNA 11.8.09	Default	11/08/2009	13:09	373.01	9.325	4.487	2.08

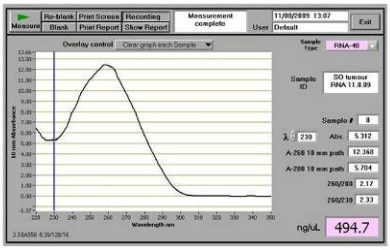


Figure 57. Examples of the Nanodrop ND-1000 spectrophotometer results for RNA isolated from fresh snap frozen surgical resection material. The table on the left shows high concentrations and good quality RNA as assessed by absorbance at 260nm and 260/280 ratios near 2.0. One sample in this batch (MF) had no identifiable RNA which was likely lost during purification and was discarded and RNA isolation repeated. The graph on the right shows an example of a clean peak at 260nm as expected for pure nucleic acid.

The RT negative controls did demonstrate some amplification by qPCR but these were at threshold levels approximately x20 higher for the very highly expressed 18S or at Ct values over 35 for GAPDH, CD45 and CK19 (Figure 46). These levels are considered to be insignificant or negative and demonstrate a lack of significant genomic DNA contamination in the purified RNA from the surgical resection material.

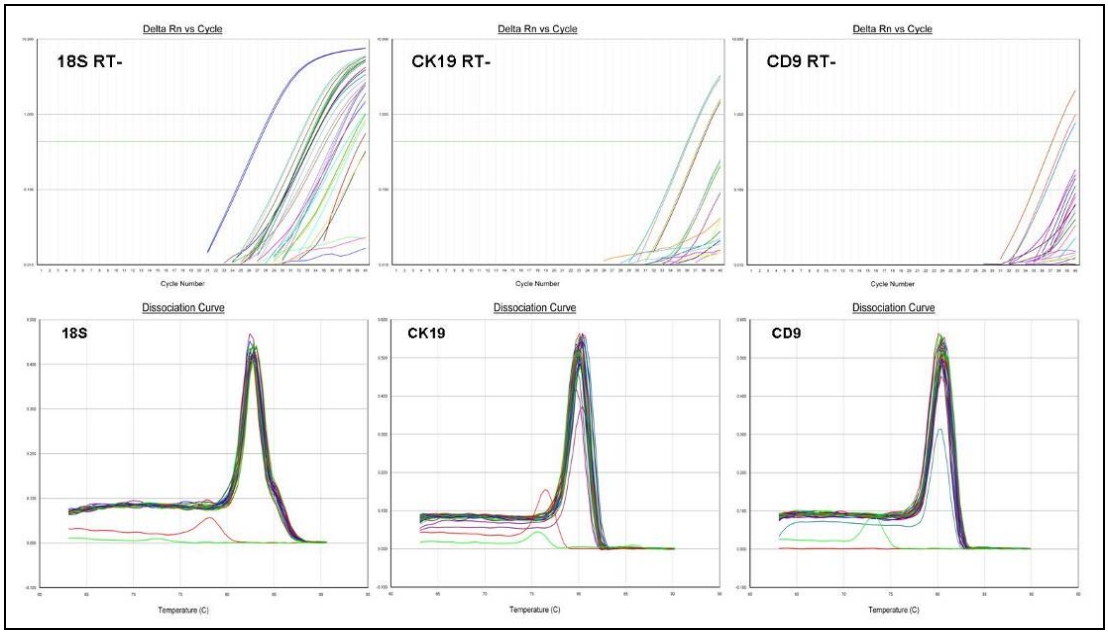


Figure 58. Reverse transcriptase negative controls for cDNA used in the qPCR experiments. Samples primed for 18S rRNA showed amplification from Ct 27 to >40 compared to approximately Ct 8 for RT positive samples (not shown). All samples for CD45 and CK19 had Ct value ≥35. These results are considered to be negative demonstrating either a lack of significant genomic DNA contamination in the source RNA and/or appropriate mRNA specific amplification of the primer sets used. Single peak dissociation plots demonstrate specific qPCR amplification and lack of non-specific signal caused by primer dimerisation.

7.5.2 Relative expression of epithelial and leukocyte RNA in surgical resection material

As per the experiments using RNA isolated from biliary brushings, we used the epithelial and pan-leukocyte markers CK19 and CD45 respectively, in order to assess the relative expression of epithelial and leukocyte RNA in benign and malignant tissue samples (n=14).

Expression of CK19 was similar in both groups with a fold change of 0.89 (95% CI 0.87 – 0.92) in cancer versus benign suggesting similar, or slightly lower but comparable, relative quantities of epithelial cells in malignant compared to benign samples.

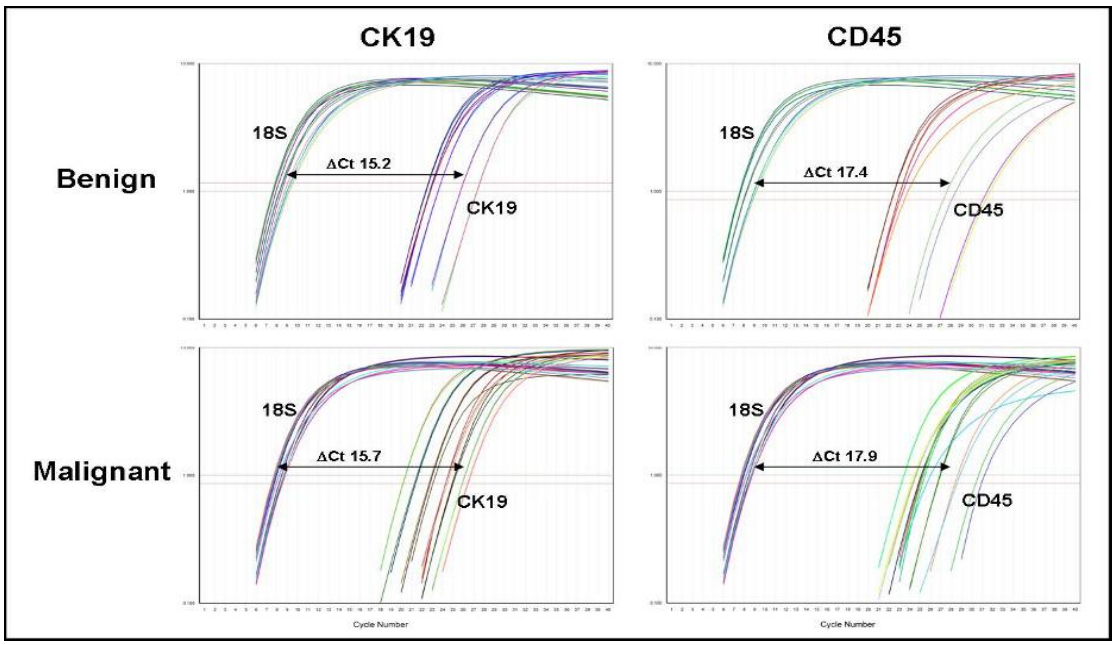


Figure 59. Examples of SYBR Green qPCR amplification plots for epithelial and leukocyte markers in benign and malignant biliary surgical resection tissues. Note similar Ct values for CK19 and CD45 in benign and malignant groups but that CD45 has significantly higher Ct values than CK19 suggesting lower levels of leukocyte than epithelial RNA.

When comparing relative expression of leukocyte RNA using CD45, there were similar contributions of leukocyte RNA in both benign and malignant samples (DCt 17.4 and 17.9 respectively). However, when comparing the relative contribution of leukocyte RNA in brushes and tissues, there was a much higher contribution of leukocyte RNA in surgical resection material (DCt 17.4) compared to ERCP biliary brushings (DCt 21.1) suggesting that biliary

brushings are a purer source of epithelial cells than surgical resection material (Figure 59).

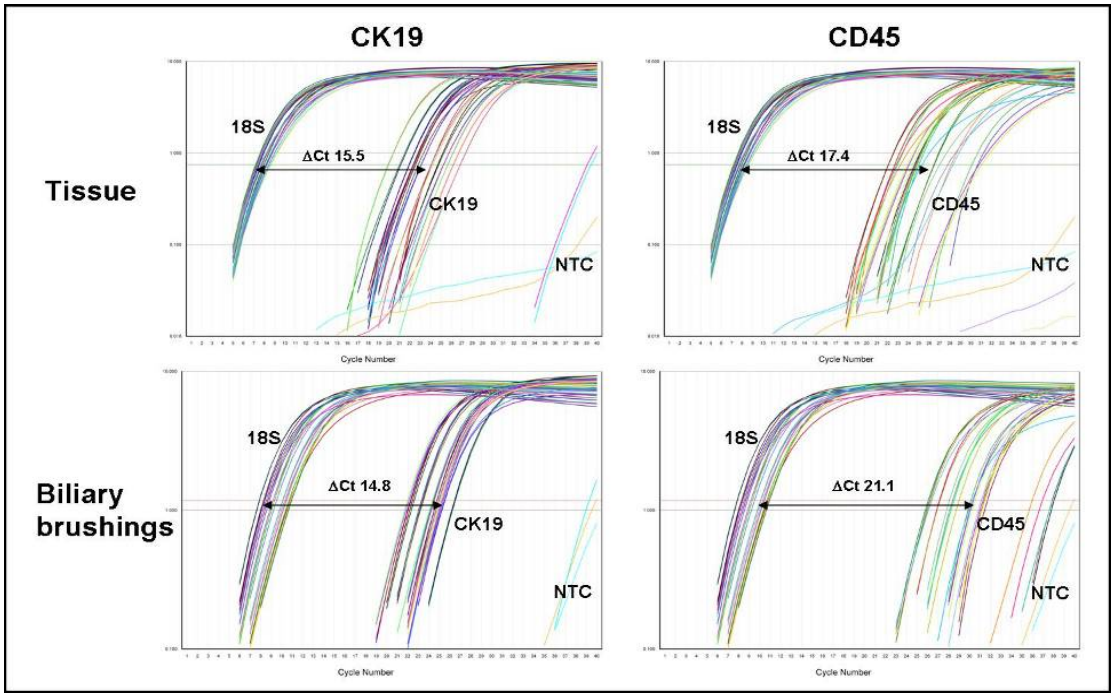


Figure 60. qPCR plots demonstrating differences in relative quantities of epithelial and leukocyte RNA from biliary tissues and biliary brushes. Note the mean ΔCt for CK19 is slightly lower in the brush samples suggesting slightly higher relative quantities of epithelial RNA in brushes than tissues. Similarly, the Ct values for CD45 were much higher in biliary brushes (21.1 v 17.4) suggesting much fewer leukocytes in biliary brush samples.

7.5.3 Analysis of multiple gene expression in BTC surgical resection material by TaqMan Array qPCR

46 genes were assessed using 2 reference genes (GAPDH and 18S rRNA) for calculation of comparative gene expression by the $\Delta\Delta$Ct method. Relative expression of each gene was calculated using Microsoft Excel (data shown in the Appendix). The results are presented in Table 26 with comparison to fold change in the biliary brush samples as calculated by microarray and qPCR. Examples of the qPCR plots are shown in Figure 61 in order to demonstrate the spread of amplification thresholds for the individual samples.

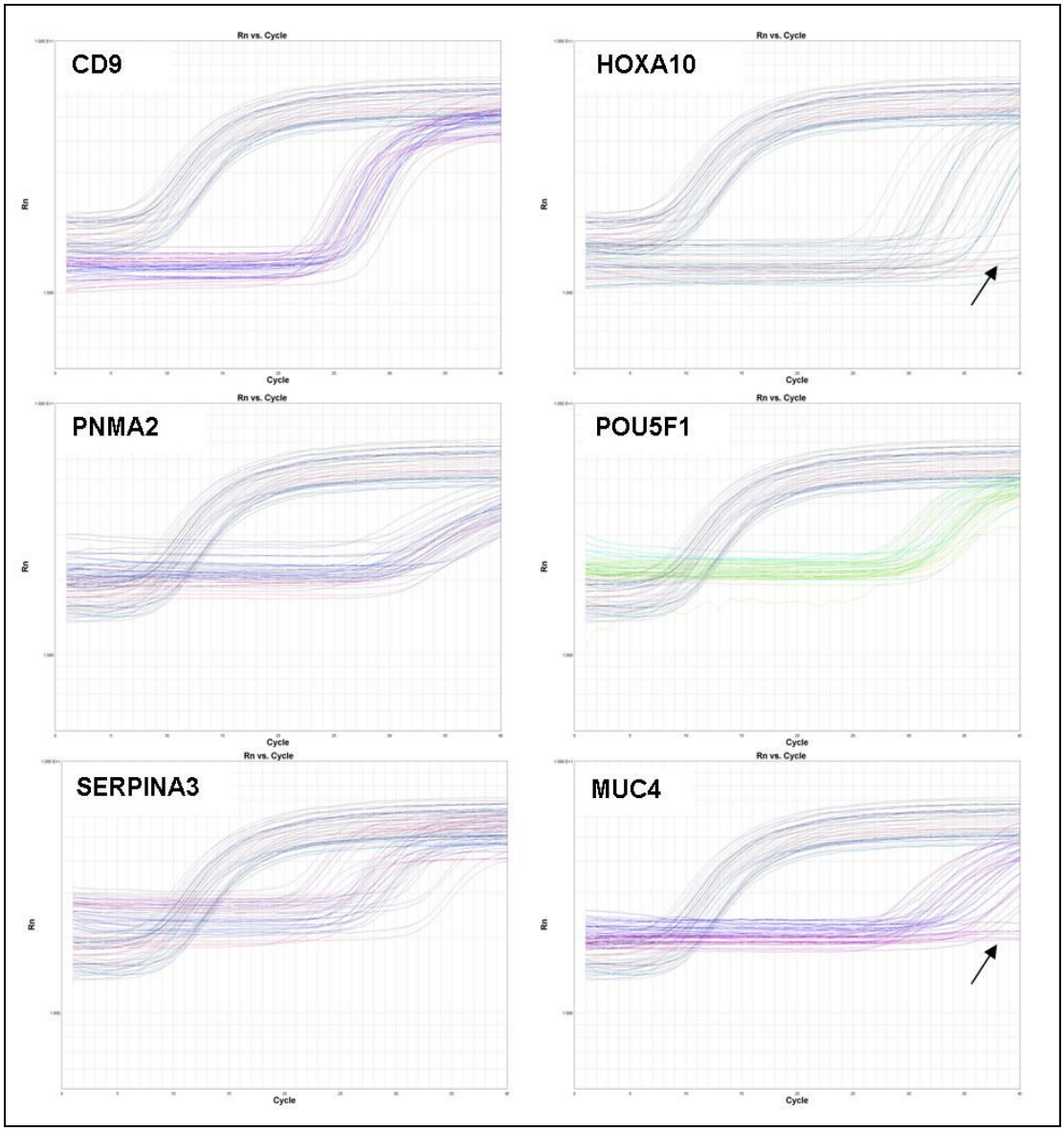


Figure 61. qPCR plots for benign and malignant biliary tissues. The pooled plots on the left of each graph are the 18S reference gene and the amplification plots for genes of interest are shown on the right hand side of each graph to show the range of cycle thresholds. Note that for HOXA10 and MUC4, some of the samples had no amplication (arrows) (ie no expression) and all of these were benign.

Table 26. Comparison of relative gene expression in cancer between biliary brushings and biliary tissue. N/A denotes genes not assessed in the TaqMan array cards for the biliary brushings or biliary tissue batches.

	Mean fold change (cancer versus benign) using RNA isolated from biliary brushings			Mean fold change using RNA isolated from biliary Tissue
Gene	Microarray	qPCR	qPCR validation set	**qPCR Tissue**
18S	1	1.0	1.0	**1**
GAPDH	1	1.0	1.9	**0.9**
ACTB	1	2.0	2.3	**N/A**
MUC4	3.2	29.4	15.5	**4.1**
MUC5AC	3.6	7.4	5.1	**1.2**
CD9	19	2.2	4.0	**1.5**
NOTCH3	6	8.0	4.7	**1.7**
ASPHD1	10	1.7	3.6	**1.0**
ATP6V0A2	10	1.1	3.1	**1.0**
CEACAM1	3.5	2.1	3.1	**0.6**
CELSR1	5.2	7.2	7.3	**1.6**
CFDP1	5.3	1.1	1.7	**0.5**
COL17A1	16	10.9	4.9	**15.6**
COL1A1	7.2	21.6	3.4	**2.5**
COL6A3	8.8	26.7	2.3	**1.6**
CSPG4	8.5	19.5	1.6	**1.3**
HOXA10	24	67.0	35.8	**6.2**
HOXB6	20	5.0	10.2	**1.5**
ITGB8	2.7	4.0	2.6	**2.0**
ITIH5	20	12.4	1.5	**0.3**
KRAS	-2.5	1.1	1.5	**0.8**
LAMC2	2.5	2.8	3.3	**1.4**
LEF1	15	7.1	4.9	**1.6**
LIF	4.7	7.5	2.0	**3.1**
LUM	11	3.7	1.1	**0.6**
MAML2	3.5	3.9	1.4	**0.8**
MAML3	-8.5	-1.3	3.7	**N/A**
MAPK1	2.35	1.6	2.2	**0.8**
MCM4	2.2	1.5	2.5	**1**
MMP2	2.6	11.0	2.3	**0.9**
MXRA5	24	10.4	3.4	**1.5**
MYC	2.7	6.0	11.5	**0.9**
NCSTN	3.4	1.0	1.9	**0.8**
PEAR1	10.3	6.6	2.7	**0.8**
PNMA2	39	3.9	112.3	**0.9**
POU5F1	3.3	2.5	5.5	**1.4**
PRKCB1	8.3	6.8	3.1	**0.5**
PTK2	3	1.9	1.8	**0.8**
PVT1	11	12.1	12.6	**5.2**
RAB27A	2.5	2.7	2.7	**0.6**
RAD51	4.3	3.0	4.2	**1.5**
SERPINA3	37.8	46.7	4.4	**0.1**
SPARC	7.1	41.4	3.1	**1.1**
STAT1	2.5	1.9	2.8	**0.8**
TGFBI	12.3	3.3	2.5	**1.2**
TM4SF18	10.9	0.9	5.3	**0.6**
TRIB2	4.3	3.4	3.1	**1.0**
VEGFA	2.5	2.6	2.9	**1.0**
MLL4	7.7	N/A	N/A	**0.9**
VIM	2.0	N/A	N/A	**0.8**

When compared to the original biliary brush qPCR data, the results of TaqMan Array qPCR in tissue had a much lower concordance than may have been expected. For the biliary brush samples, the pattern of gene expression was similar in both the microarray and qPCR data with 79% (34/43) concordance of overall gene expression between the 2 platforms when comparing up-regulated genes. The relative gene expression (fold change) varied but the trend was considered similar if over 2 in both the microarray and qPCR data. Using the second, fresh validation set, 36 out of 43 upregulated genes showed a similar pattern in gene expression (83% concordance).

In the tissue samples, using the same criteria (cut-off of ≥2) for upregulated genes, the concordance between biliary brush and tissue mRNA expression was much lower at 16% (7/43). Genes that were consistently elevated between microarray and qPCR of biliary brush samples with similar elevated expression in the tissue samples were MUC4, COL17A1, COL1A1, HOXA10, ITGB8, LIF, and PVT1. Other genes were not considered significantly upregulated in the tissue samples and others appeared downregulated in tissue despite upregulation being shown in the biliary brush samples (CEACAM1, PRKCB1, RAB27A, SERPINA3, and TM4SF18).

7.5.4 Relative expression of CD9 assessed by single gene SYBR Green qPCR

Samples used for single gene qPCR were the same as those used for the TaqMan Array qPCR except for the loss of one of the benign samples due to insufficient RNA/cDNA remaining (benign n=5, BTC n=9). The ΔΔCt method was used to calculate the expression of each gene relative to the reference gene (usually 18S rRNA). The relative expression of CD9 was higher in cancer versus benign samples (fold change 2.38 [95 % CI 2.24-2.53], Figure 62) despite the slightly lower epithelial content suggesting an underestimate of fold change in cancer.

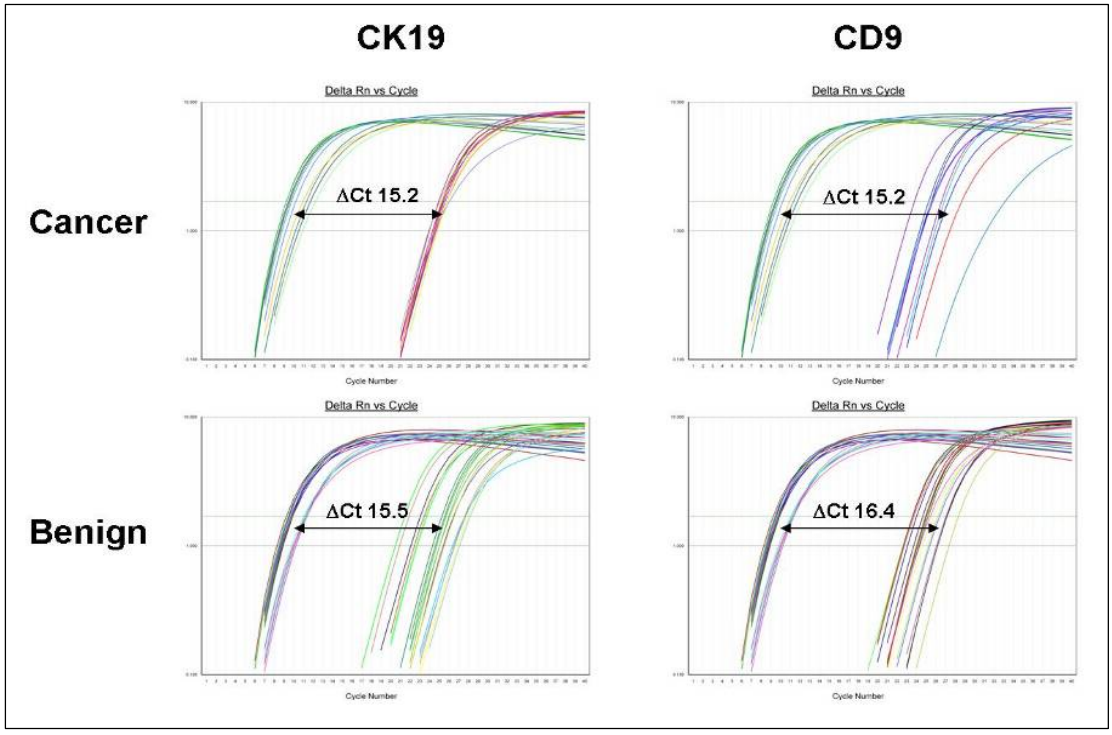


Figure 62. SYBR Green qPCR using RNA isolated from fresh frozen surgical resection material. The similar ΔCt values for CK19 in the benign and malignant samples suggest similar quantities of epithelial tissue but the lower ΔCt values demonstrate higher CD9 levels in cancer samples.

For one patient undergoing a resection of colorectal liver metastasis, we had separate samples of a portion of normal liver as well as the dissected left main bile duct. RNA was isolated from both tissues and analysed separately to assess relative expression of CK19 and CD9 (Figure 63). As might be expected, CK19 mRNA expression was higher in the dissected bile duct tissue than in the liver section (ΔCt 15.9 and 18.1 respectively). The expression of CD9 mRNA however was higher in the normal liver than in biliary tissue (14.7 and 17.0 respectively).

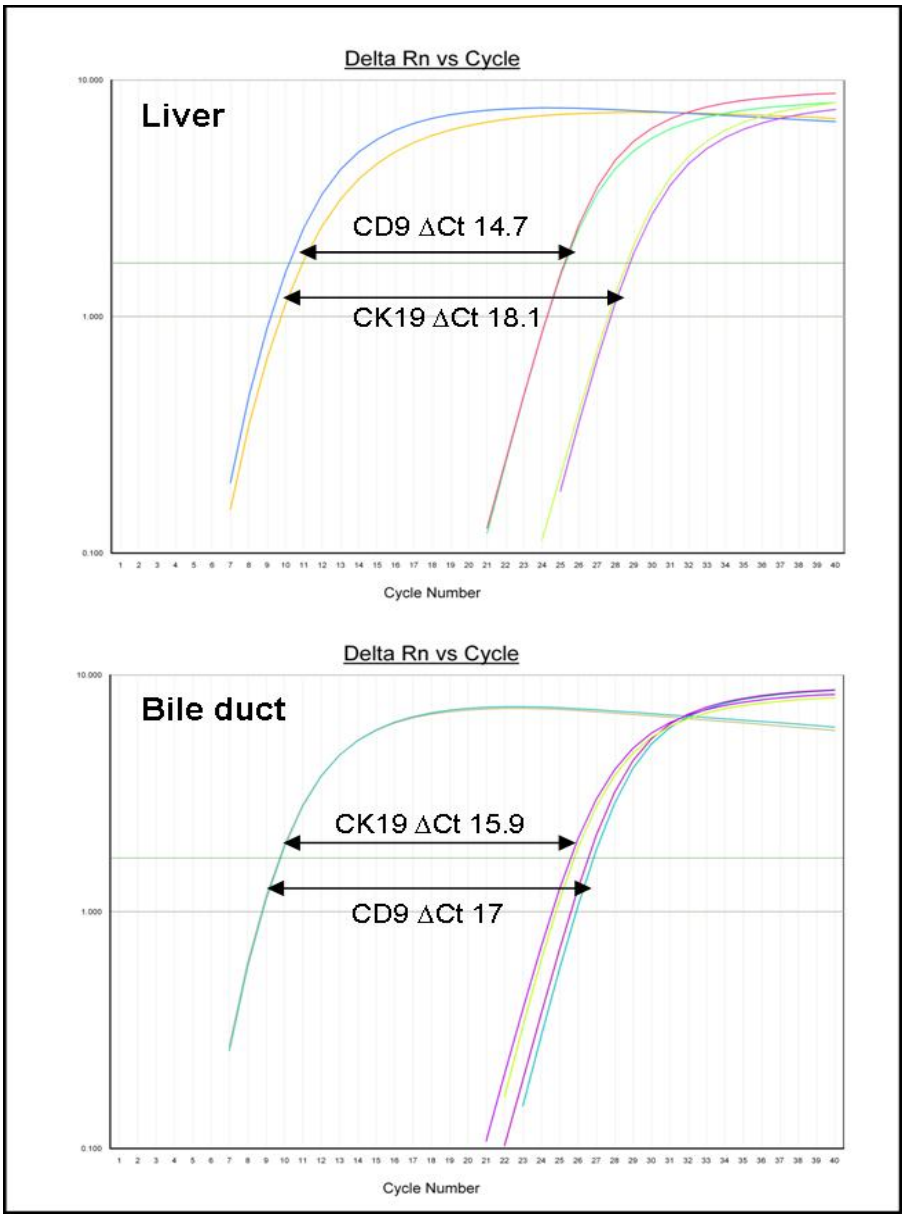


Figure 63 Relative expression of CK19 and CD9 in a sample of normal resected liver compared to normal left main bile duct showing higher expression of CK19 and lower expression of CD9 in biliary tissues compared to normal liver.

7.6 Conclusions: analysis of mRNA expression in biliary surgical resection material.

These data demonstrate that measurement of mRNA expression can be done by multiple gene qPCR as supported by the up-regulation of MUC4 as previously described in BTC. However, despite the better quality RNA, the relative expression of epithelial and leukocyte RNA suggest that complex, homogenised whole fragments of biliary tissues are less suitable for analysis of epithelial cell gene expression than those obtained from ERCP biliary brushings. The high stromal component of BTC tissue is well recognised and examples of the mix of cell types in BTC tissue is shown in the immunohistochemistry figures later in this chapter. Possibly for these and other reasons, when compared with gene expression in ERCP biliary brushings, elevated mRNA expression in surgical resection material had a concordance of only 16%. However, a selection of genes with consistently elevated mRNA expression in all sample sets have been identified and deserve further investigation. These include PVT1, HOXA10, COL17A1 and POU5F1 (oct 3/4).

7.7 Part 2. Analysis of protein expression of genes identified as elevated in biliary brushings.

7.8 Introduction

We have demonstrated elevated expression of a number of genes in biliary brushings and surgical resection material. However, elevation of gene expression may not always translate to elevated protein expression due to non coding gene mutations in the genes of interest, post translational modification of mRNA by microRNA and other means, or rapid degradation of protein products. Some proteins may be identifiable in serum but would more likely be present in bile or biliary cancer tissue. The most commonly used methods for investigation of protein expression are Western blot, ELISA and immunohistochemistry.

Western blot uses radio- or chemiluminescent-labeled antibodies directed towards epitopes on the proteins of interest in order to semi-quantitatively assess levels of protein expression in tissues. There are a number of limitations with this method, which is difficult to apply to complex and highly variable biological fluids such as bile. Quantification of protein expression by Western blot is also very limited and is primarily used to demonstrate presence or absence of protein. Protein expression analysis in biological fluids by enzyme linked immunosorbance assay (ELISA) is generally much

simpler to perform, more reproducible and is quantitative. However, ELISA kits and/or ELISA tested antibodies were not available for many of the proteins of interest leaving Western blot as the primary method for protein identification in our clinical samples.

Immunohistochemistry is a standard and reliable method for assessing protein expression in tissue and was chosen as the primary method for further investigation. Immunocytochemistry of biliary brush cytology specimen may provide a more direct comparison of the biliary brush gene expression data but cytological samples usually provided only 1-4 suitable slides and were not practical for confirmation of multiple protein expression. Bile is simpler to collect but is a complex biological fluid and much less simple to obtain reliable measurement of protein expression. Also, bile is often diluted to variable amounts by x-ray contrast agents used at ERCP, and there are no clear reference protein standards with consituents of bile varying in health and disease and in different anatomical locations within the biliary tree. The level of common protein standards such as albumin and cellular protein standards such as B-actin vary with disease and the cellularity of bile.

7.9 Part A: Analysis of protein expression in bile.

If elevated gene expression in epithelial cells in the lumen of bile ducts is translated to elevated protein expression, then bile is likely to be the body fluid with the most potential for having associated elevated protein levels in biliary diseases such as BTC. As we showed previously with MUC4 (Matull, Andreola et al. 2008), measurement of protein expression in bile can be performed by Western blot using suitable conditions and antibodies. We aimed to assess the expression of a number of proteins not previously analysed in bile.

7.10 Materials and methods

7.10.1 Western Blotting

7.10.1.1 Clinical samples

Samples of bile obtained at the time of ERCP or from percutaneous external biliary drains were used for measurement of protein by Western blot. Samples were obtained followiing ethical approval (NHNN ethics committee, ref 06/Q0152/106) and informed patient consent. All bile samples were stored at -80^{0}C. Total protein isolated from human BTC cells lines (TFK-1 and HUCCT) was used for positive controls. For some experiments, bile was filtered using MicroCon YM-100 Centrifugal Filter Devices (Millipore) to remove large, complex biliary glycoproteins that may have interfered with the gel separation and/or antibody binding.

7.10.1.2 Gel electrophoresis separation of bile proteins.

Bile is a highly heterogeneous biological fluid and samples had often been diluted by x-ray contrast agent used at the time of ERCP. Therefore, total protein in bile was measured using a standard bicinchoninic acid (BCA) method in order to calculate the volume of bile required to obtain a total protein quantity of 30µg suitable for Western blot analysis. Protein concentration was measured in 96 well plates using the MicroBCA Protein Assay Kit (Pierce Protein Research Products, Rockford, IL, USA) with a standard curve of bovine serum albumin and an ELISA plate reader used to calculate protein levels. Bile samples were made up to a total volume of 25µl with water and 1.25µl of 1M Sample Reducing Agent (1M DTT, Invitrogen, Paisley, UK) followed by heating to 70^0C for 10mins for denaturing of proteins. Gel electrophoresis separation was carried out at 200V using 4-12% NuPAGE® Bis-Tris gels (Invitrogen) in 1% NuPAGE MOPS SDS Running Buffer (Invitrogen) alongside 10µl of SeeBlue® Plus 2 Pre-stained Protein Standard ladder (Invitrogen).

7.10.1.3 Western Blotting protocol.

A summary of the steps used for a standard protocol for Western blotting were as follows:

1. Gel electrophoresis separated proteins were transferred to Invitrolon PVDF Filter paper sandwiches (Invitrogen) using an iBlot® Dry Blotting

System (Invitrogen) as per the manufacturers instructions. Membranes were then:

2. stained with Ponceau red to ensure adequate protein transfer and washed with distilled water.
3. blocked using 0.5x Western Blocking Reagent (Roche) for 1 hour and incubated at 4^0C overnight with 10ml of primary antibody solution, usually at a dilution of 1:1000 in 0.5x Western Blocking Reagent in TRIS buffered saline with Tween (TBS.T).
4. washed x2 with TBS.T (0.1% concentration Tween) followed by x2 washes with 0.5x Western Blocking Reagent in TBS-T (10 mins each).
5. incubated with 10ml of secondary antibody solution for 1 hour at room temperature at a dilution of 1:5000
6. washed x2 with TBS-T followed by x2 washes with 0.5x Western Blocking Reagent in TBS.T (10mins each).
7. incubated with 5ml chemoluminescence substrate (SuperSignal® West Pico, mouse or goat horseradish peroxidise kits, Thermo Scientific) for 5 minutes.
8. exposed onto plain photographic film for 1, 2, 5, 15 and 30 mins until visible bands were seen on developing the film.

Western blot experiements were done with the assistance of Robert Tidswell (iBSc student) and guidance of Dr Virginie Cerec and Dr Fausto Andreola.

7.10.1.4 Antibodies used for Western blots

Primary antibodies

1. α-tubulin, mouse anti-human IgG monoclonal antibody (Upstate, #05-829, clone DMIA) was used as a reference protein, as the small 55kDa protein product is ubiquitously expressed in the cytoskeleton of all cells.
2. Notch3, rabbit anti-human IgG (Cell Signalling, #2889) recognising full length 270 KDa and extracellular 90 KDa fragments of the notch3 protein. Notch 3 was chosen for Western blot as reliable tested antibodies were available in the laboratory and the gene was reported to be 4.7 to 8 fold up-regulated in malignant biliary brushings by qPCR and microarray. There was a lesser up-regulation of mRNA (1.7) in the BTC tissue samples.

Secondary antibodies

1. Affinity purified horse anti-mouse, horseradish peroxidise linked antibody (#7076, Cell Signalling Technology)
2. Affinity purified goat anti-rabbit, horseradish peroxidise linked antibody (#7074, Cell Signalling Technology)

7.11 Results

7.11.1 Measurement of α-tubulin protein levels in bile.

Calculated total protein levels in bile were highly variable with a large range from 0.8g/ml to 20g/ml. (Example, Table 27).

Table 27 Examples of total protein calculations for bile by ELISA using a standard curve of bovine serum albumin. Note the highly variable protein concentration likely related to presence of confounding factors such as cholangitis, biliary obstruction and source of bile (eg gallbladder, ERCP, or internal/external drainage tube).

SAMPLE ID	Crude Concentration μg/ml
Control; saliva	7268.8
Bile 33	4824.4
Bile 19	874.78
Bile 14	19360
Bile 28	5097
Bile 29	16541.6
Bile 29 supernatant	3612.2
Bile 21	34006
Bile 26	4511.2
Bile 16	4268.8
Bile 46	16844.8
Bile 60	3733.4
Filtered bile 34	1425.4
Pure bile 34	4268
Filtered bile 50	959.92
Pure bile 50	2940

On Western blot, no 51KDa protein bands were seen in any of the bile samples for up to 30mins exposure despite clear bands being visible in the TFK cell protein extracts (Figure 65). These results may be explained by:

1. Total protein measurements in bile include a large proportion of extracellular and secreted biliary proteins not found in the TFK-1 cells and there may have been a relatively small amount of cellular protein material that was insufficient to produce visible a-tubulin bands in the Western blot gel.
2. Complex biliary glycoprotein fragments separated in the gel may inhibit protein transfer or binding of antibody to the α-tubulin protein on the membrane.
3. The large complex biliary glycoproteins will not likely separate out on the 4-12% gels used and the smaller α-tubulin protein may have been trapped within the complex biliary glycoproteins producing the bands seen near the position of the loading wells (Figure 65).

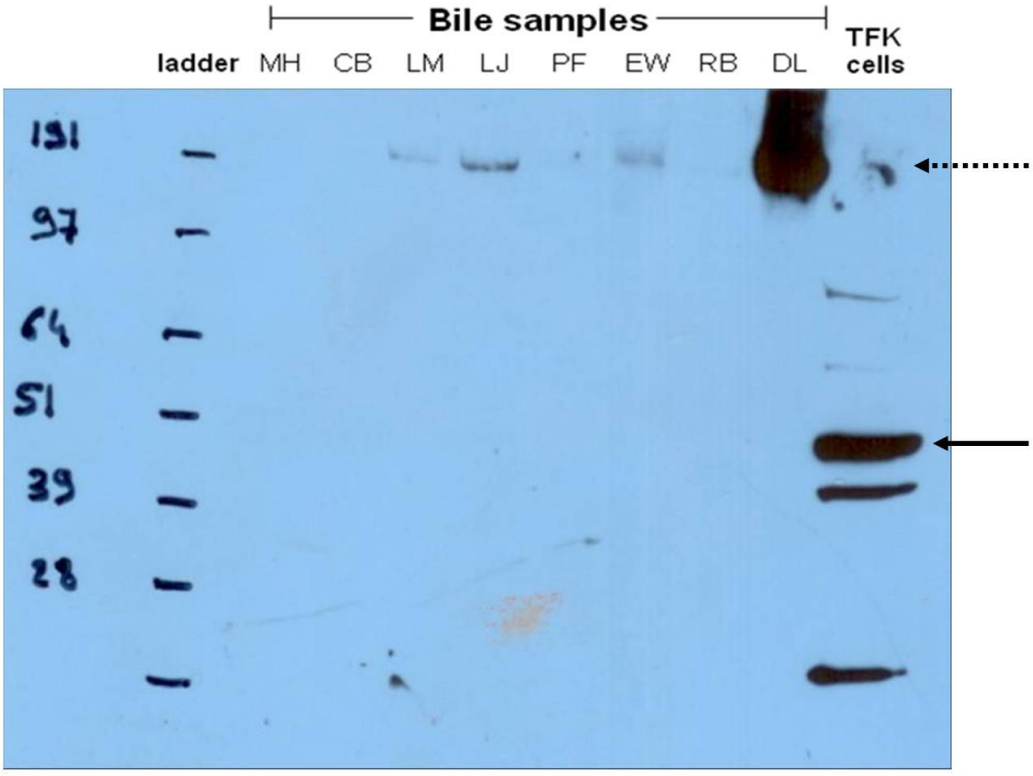


Figure 64. Western blot of bile samples using antibodies directed against a 51KDa alpha-tubulin protein showing clear bands in the TFK-1 controls (black arrow, smaller fragments are also idenfied by the antibody) but bands of variable intensity in the large molecular weight areas in the bile samples (dashed arrow) suggesting evidence of impaired gel electrophoresis in the bile samples.

When repeated with filtered and unfiltered bile from the same patients, results showed mixed levels of staining suggesting highly variable and unreliable protein filtration or transfer with possible reasons outlined above.

When using known and predictable quantities of cellular protein as a pure extract or spiked into bile, the size of the bands was markedly reduced in the protein spiked in bile, suggesting impaired gel electrophoresis, transfer or antibody binding as described above (Figure 66).

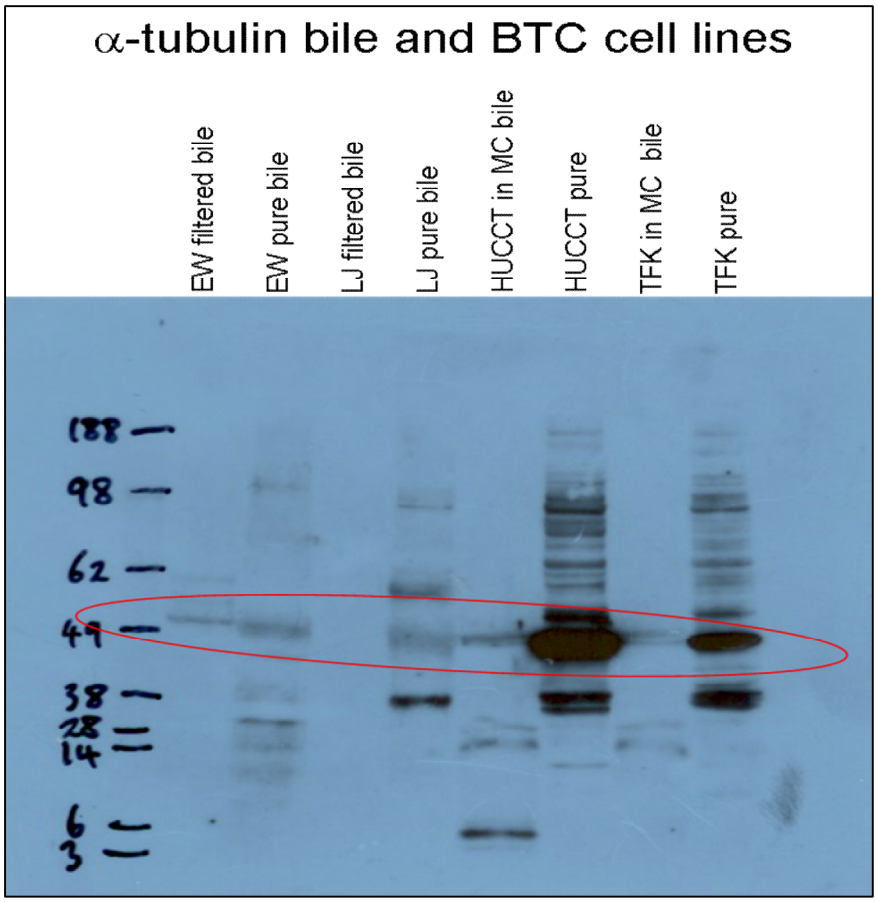


Figure 65. Results of Western blot for α-tubulin in bile samples with or without filtering of large molecular weight proteins. Note that with the BTC cell line controls, cells spiked in bile resulted in a significant reduction in the blot size suggesting that bile impairs the process and applicability of Western blot analysis.

7.11.2 Measurement of notch3 protein in bile

When notch3 protein expression assessed by Western blot was performed using bile samples, no visible bands of notch3 protein were found suggesting an absence of notch3 protein or methodological problems as outlined above for α-tubulin (data not shown).

7.12 Conclusions: Western blot analysis of protein expression in bile.

As cholangiocytes line the lumen of bile ducts and these are the primary source of RNA providing data on gene expression in the biliary brushings, bile would appear to be a good potential source of tissue for analysis of protein expression and development of diagnostic tests.

Despite previously published works describing alterations in biliary proteins in disease (Matull, Andreola et al. 2008), these results suggest that Western blot is not a reliable method for investigating the presence of proteins of interest in bile. Total protein levels in bile are highly variable and the composition of protein types (e.g. cellular proteins, extra-cellular glycoproteins, immunoglobulins etc) is also highly variable and not easily to quantitate. In view of this and the variable cellularity of bile, there is no definitive reference protein standard for use with bile samples making quantification by Western blot difficult. In addition, other than the dilutional effects of x-ray contrast agents in bile, it is unclear how these agents may interfere with the methodology of Western blot. Our data also suggest major interference in the assay by components of bile which may include the large complex glycoproteins. Previously published work used a higher degree of bile filtration with long (many hours) centrifugation times through small pore filter columns to exclude larger biliary proteins. It is unclear whether this process efficiently separates small proteins or protein fragments of interest and so quantification of proteins in the filtrate may also be unreliable.

In view of the methodological difficulties and lack of reliable quantification using these methods, protein expression would be better assessed in bile using alternative methods such ELISA or assessed in tissue sections using immunohistochemistry. The lack of commercially available ELISA kits for our proteins of interest supported the use of immunohistochemistry as the most suitable next step in analysis of protein expression.

7.13 Part B: Assessment of protein expression in biliary tissues using immunohistochemistry.

7.14 Introduction

Immunohistochemistry (IHC) is a standard method for assessing protein expression in histology sections of tissue specimens. It is used both in research and for clinical assessment of patient samples. The methods require standardised protocols for each tissue type with different conditions for antigen unmasking and retrieval, antibody binding and staining.

In order to further assess protein expression of a selection of genes of interest, we aimed to use IHC to assess distribution and levels of staining of some of these proteins. One gene that appears to be highly upregulated in BTC (PVT1) is not translated to protein so cannot be further assessed by this method. Its mechanisms of action are unclear but include the coding of micro RNA sequences that alter mRNA expression of other genes.

7.15 Materials and methods

7.15.1 Samples

Tissue blocks stored at the UCL Department of Pathology from patients who had undergone surgery for benign and malignant disease were identified from clinical databases. Ethical approval for use of the tissue was granted as

described previously (06/Q0152/106). Cutting of blocks and immunohistochemistry work was performed as a paid service by UCL Advanced Diagnostics (Philippa Munson, R&D Manager) as UCL regulations do not allow tissue blocks of clinical samples to be released from the Department of Pathology. A summary of the clinical samples used for immunohistochemistry is shown in Table 28.

Table 28. Summary of patient details of sections used for immunohistochemistry

	Sample	Diagnosis	Age	Gender
Benign	**B1**	CBD (Choledochal cyst)	74	F
	B2	GB (Chronic cholecystitis)	63	F
	B3	GB (Chronic cholecystitis)	44	M
	B4	GB (Chronic cholecystitis)	30	F
	B5	GB (Chronic cholecystitis)	39	F
	B6	GB (Chronic cholecystitis)	15	F
	B7	Gb (Chronic cholecystitis)	43	F
	B8	Cystic duct (Chronic cholecystitis)	40	F
	B9	GB(Chronic cholecystitis)	52	F
	B10	GB (Chronic cholecystitis)	57	M
	B11	CBD (Chronic pancreatitis)	50	M
	B12	(Chronic cholecystitis)	67	M
			Mean 48	
BTC	**C1**	Tumour mass (CC)	55	M
	C2	Tumour mass (GBCa)	63	F
	C3	Tumour mass (CC)	65	M
	C4	Tumour mass (CC)	59	M
	C5	Tumour mass (CC)	75	F
	C6	Tumour mass (CC)	53	M
	C7	Tumour mass (CC)	69	M
	C8	Tumour mass (CC)	49	M
	C9	Tumour mass (CC)	55	M
	C10	Tumour mass (GBCa)	78	M
	C11	Tumour mass (CC)	61	M
	C12	Tumour mass (CC)	56	F
			Mean 62	

7.15.2 Selection of proteins for analysis by immunohistochemistry.

The proteins for further analysis were partly selected prior to completion of the tissue mRNA expression analysis. Criteria used to select proteins of interest included;

1. Genes consistently up-regulated in different validation sets
2. Proteins with antibodies suitable for IHC using FFPE tissues (determined by either manufacturer datasheets or other published work providing details on suitable methods).
3. Proteins already identified as biomarkers in other cancers
4. Proteins expressed in extracellular compartments or those that may be secreted into body fluids (bile or blood).

Some candidate proteins did not fulfil sufficient of the criteria above and were not further assessed by IHC. Proteins selected for analysis by IHC and details of suitable antibodies used are shown in Table 29.

Table 29. Summary of proteins assessed and antbodies used for immunohistochemistry

Protein	Antibodies	Positive control	References
HOXA10	Goat anti-human HOXA10 polyclonal IgG. #SC-17159, Santa Cruz Biotechnology	Endometrium, colon	(Sarno, Kliman et al. 2005)
	Rabbit anti-human HOXA10 polyclonal IgG. #LS-C30941, Lifespan Biosciences	Endometrium, colon	Manufacturer approved for FFPE tissues
POU5F1 (Oct3/4)	NCL-L-OCT-3/4, Novocastra Laboratories, Burlington, Canada)	seminoma	In routine clinical use by the UCL pathology lab
PNMA2	#sc-68098 (C-19), Santa Cruz	Neuronal cells	(Leja, Essaghir et al. 2009)
	Rabbit anti-human IgG #HPA001936, (Atlas Antibodies (Sigma))	Neuronal cells	Manufacturer approved for FFPE tissues
SERPINA 3 (AACT)	Mouse anti-human SERPINA3 monoclonal IgG M02 (clone 1C10) # H00000012-M02, Abnova	Colon, skin	Manufacturer approved for FFPE tissues
CD9	Monoclonal IgG mouse anti-human CD9. #NCL-CD9, Novocastra Laboratories, Burlington, Canada)	Mammary fibroadenoma	(Coudry, Meireles et al. 2007)
COL17A1	Mouse anti-human COL17A1 monoclonal IgG, #NC16A-3, (ab79878), Abcam	skin	Manufacturer approved for FFPE tissues

7.15.3 Immunohistochemistry methods

This work was performed as a paid service by UCL Advanced Diagnostics using the Bond™ Automated Immunohistochemistry & In-Situ Hybridisation System. Tissue stored as FFPE blocks were serially cut to 10μm sections for IHC. One section was stained with Haematoxylin and Eosin (H&E) for later assessment of tissue sections by a histopathologist.

Tissue sections were mostly processed with similar steps apart from minor differences in incubation times, enzymes and secondary antibodies etc. A summary of the steps for IHC is as follows;

1. Deparaffinisation using x3 washes in Bond™ Dewax Solution at 70^0C followed by x3 washes in ethanol and x3 washes in Bond™ Wash Solution.
2. Antigen retrieval by incubation with Bond™ ER Solution 1 or 2 for 10 for 30mins at room temperature, 35^0C or 100^0C followed by washes in ER Solution and Bond™ Wash Solution.
3. Addition of peroxide block for 5 mins followed by x3 washes with Bond™ Wash Solution.
4. Incubation of primary antibody followed by x3 washes.
5. Incubation with the secondary antibody for 8 mins followed by x3 washes.
6. Incubations with polymer followed by x3 washes and DAB biotin substrate followed by x3 washes.
7. Staining with haematoxylin followed by x3 washes.

Variations from the main protocol outlined are described for each antibody below in line with published methodology or advice of the manufacturers.

7.15.3.1 CD9

Antigen retrieval was performed using Bond^{TM} ER Solution 1 for 30mins at 100^{0}C. The primary antibody (mouse monoclonal IgG1, NCL-CD9, Novocastra Laboratories, Burlington, Canada) was used at a dilution of 1/400.

7.15.3.2 POU5F1 (oct 3/4)

Antigen retrieval was performed using Bond^{TM} ER Solution 2 for 20mins at 35^{0}C. The primary antibody (mouse monoclonal IgG1, NCL-L-Oct3/4, clone N1NK, Leica Microsystems) was used at a dilution of 1/100.

7.15.3.3 PNMA2

Antigen retrieval was performed Bond^{TM} ER Solution 2 for 30mins at 35^{0}C. The primary antibody (rabbit polyclonal, HPA001936, (Atlas Antibodies) was used at a dilution of 1/500.

7.15.3.4 SERPINA3

Antigen retrieval was performed using Enzyme 1 for 10mins at 37^0C. The primary antibody (mouse monoclonal IgG2, H00000012-M02, Abnova) was used at a dilution of 1/400.

7.15.3.5 HOXA10

Antigen retrieval was attempted using Enzyme 1 and Bond^{TM} ER Solutions 1 and 2 for 10, 20 and 30mins. The primary antibody (LS-C30941, Lifespan Biosciences and SC-17159, Santa Cruz Biotechnology) was used at concentrations of 1/20 to 1/400 and longer incubation times. Different secondary antibodies were also tried. None of the protocols gave positive staining in even the positive endometrial tissue controls and so no suitable slides were obtained for further histopathological assessment.

7.16 Immunohistochemistry

7.16.1 Histopathological analysis of the immunohistochemistry slides.

Stained tissue sections were assessed and scored by a histopathologist, Dr Maesha Deheragoda, UCL. A standard scoring system was used to score both intensity and number of epithelial cells stained (Figure 66). The number of epithelial cells stained positive was scored as an estimate of the percentage of the total number of cells.

Figure 66. Table of criteria used for scoring of immunohistochemistry of BTC and benign tissue sections.

Level of staining	
0	None
1	low intensity
2	moderate intensity
3	high intensity

Comparison of results between benign and malignant groups was performed using a two tailed t-test.

7.16.2 Concordance with previous analysis of protein expression shown by other groups.

The protein expression pattern for genes of interest was searched in the literature and via the Human Protein Atlas database. No similar data were found in the published literature but four of the proteins of interest were reported in liver and BTC in the Human Protein Atlas database (Figure 67) (proteinatlas.org). These suggest negative or weak staining for HOXA10, CD9

and PNMA2 in normal liver and biliary epithelium and up-regulation in BTC. SERPINA3 was strongly expressed in both normal liver and BTC.

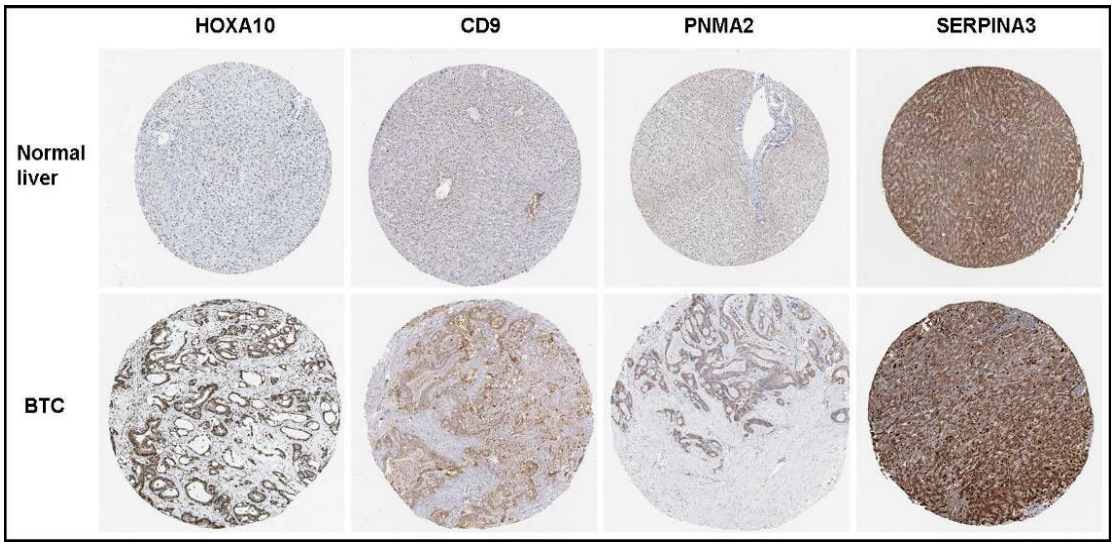


Figure 67. Examples of immunohistochemistry staining published by the Human Protein Atlas showing staining levels in normal liver (top row) and cholangiocarcinoma tissue (bottom row) for 4 of our proteins of interest. Note the increased staining in cholangiocarcinoma specimen but also heavy staining of SERPINA3 in normal liver. There were no data on either POU5F1 or COL17A1 available in the Human Protein Atlas database (proteinatlas.org).

7.17 Results: Immunohistochemistry

Results for the individual scoring and evaluation of differences between benign and malignant groups are shown in Table 30. Note that one of the benign samples is blank as it was not scored due to a lack of epithelial tissue in the histology slides.

CD9, SERPINA3 and PNMA2 were all significantly elevated at the protein level in malignant versus benign biliary epithelial tissues. Significant differences were seen both in the intensity of staining and percentage of epithelial cells staining positive (Table 30). Examples of the immunostaining for these proteins are shown in Figure 68, Figure 69, Figure 70 and Figure 71.

Despite evidence of mRNA upregulation in BTC, POU5F1 immunostaining was negative in all benign and cancer specimens. The positive control slides did however stain positive and the antibody was regularly used by the UCL Advanced Diagnostics laboratory suggesting that the immunohistochemistry results were valid and that POU5F1 (oct3/4) is not heavily expressed in biliary epithelium. HOXA10 was not assessed due to difficulties in finding suitable conditions for positive staining with the antibody. Work with HOXA10 and COL17A1 is ongoing. However, in support of our data, there is a single example of HOXA10 immunostaining staining in CC on the Human Protein Atlas online databse (proteinatlas.org) showing strong staining in CC but not in benign liver or biliary epithelium (Figure 67). There are no data available for COL17A1.

Table 30. Results of immunohistochemistry showing scoring for CD9, POU5F1, SERPINA3 and PNMA2

CD9				POU5F1			
cancer		benign		cancer		benign	
Intensity	percentage epithelial cells stained	Intensity	percentage epithelial cells stained	Intensity	percentage epithelial cells stained	Intensity	percentage epithelial cells stained
2	30			0	0		
3	30	0	0	0	0	0	0
2	20	0	0	0	0	0	0
2	20	0	0	0	0	0	0
1	20	0	0	0	0	0	0
0	0	0	0	0	0	0	0
1	10	0	0	0	0	0	0
1	30	0	0	0	0	0	0
1	5	2	30	0	0	0	0
0	0	1	20	0	0	0	0
1	40	0	0	0	0	0	0
2	70	0	0	0	0	0	0
p		0.0036	0.0112	p		NS	NS
SERPINA3				PNMA2			
cancer		benign		cancer		benign	
Intensity	percentage epithelial cells stained	Intensity	percentage epithelial cells stained	Intensity	percentage epithelial cells stained	Intensity	percentage epithelial cells stained
1	10			0	0		
2	50	0	0	2	50	0	0
2	50	0	0	2	75	0	0
1	20	0	0	1	20	0	0
1	40	0	0	1	10	0	0
3	75	0	0	0	0	0	0
3	75	0	0	1	20	1	20
2	40	0	0	2	40	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
1	10	0	0	3	80	0	0
2	80	1	10	1	80	0	0
p		0.0004	0.0012	p		0.0057	0.0098

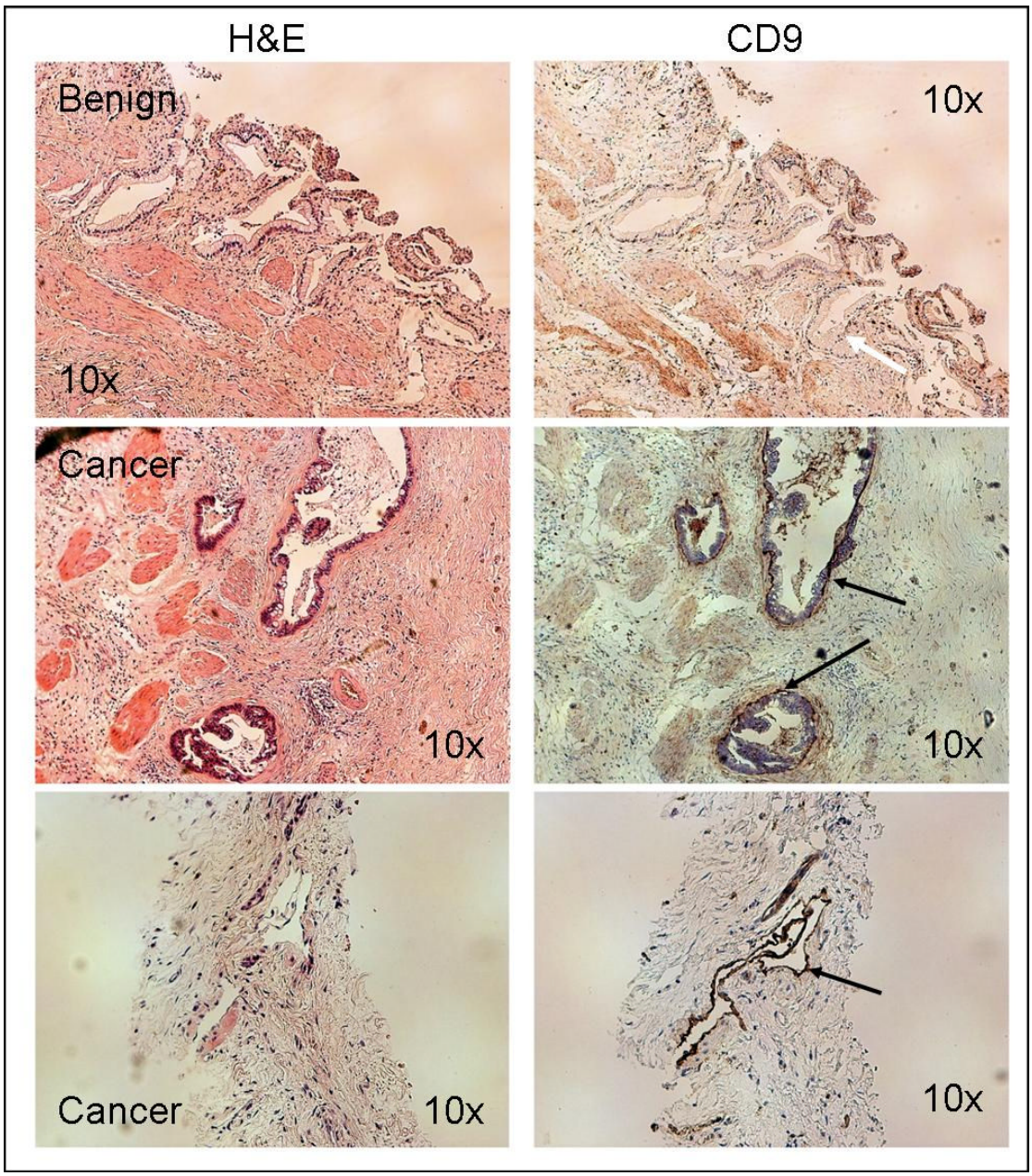


Figure 68. Examples of H&E and CD9 immunostaining for benign and malignant biliary tissues. Note the strongly positive epithelial staining for CD9 in cancer (arrows) not seen in the benign sample.

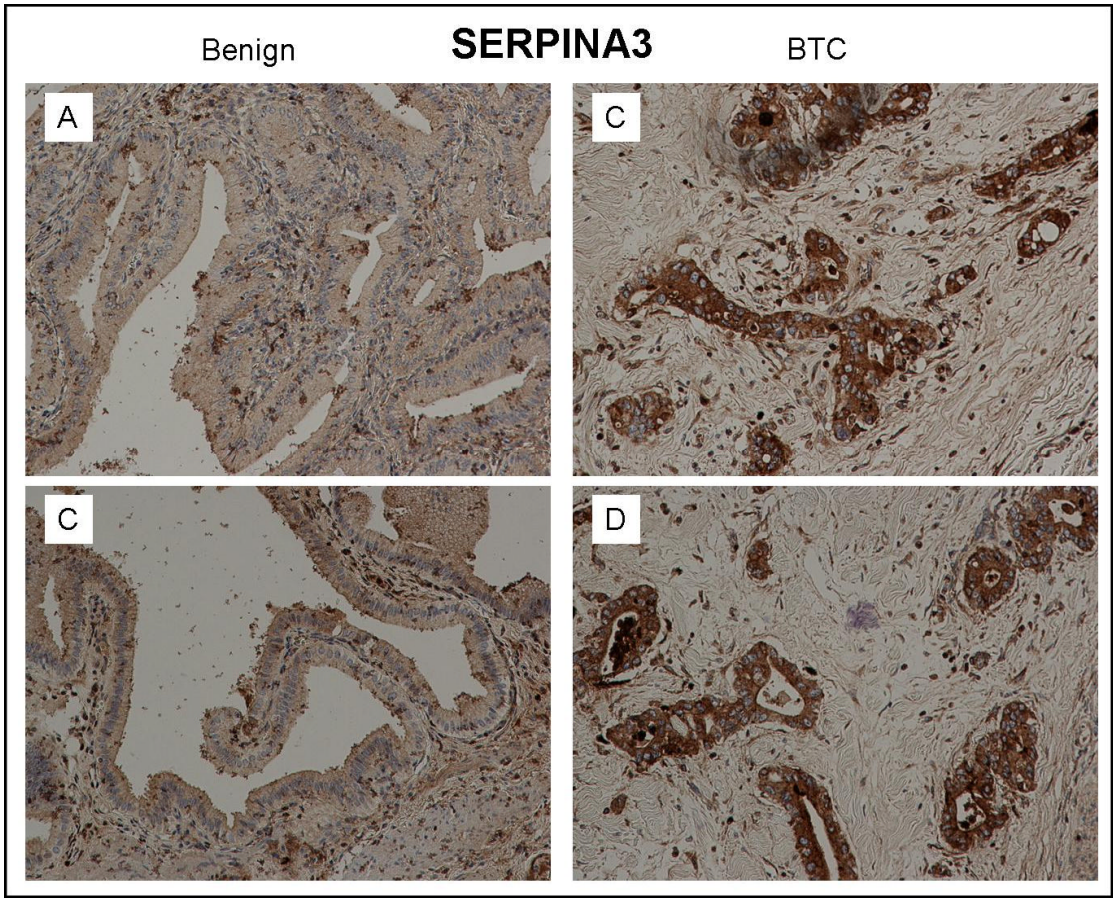


Figure 69. Examples of SERPINA3 immunostaining in benign (A&B, no staining) and malignant (C&D, strong epithelial staining) biliary tissues.

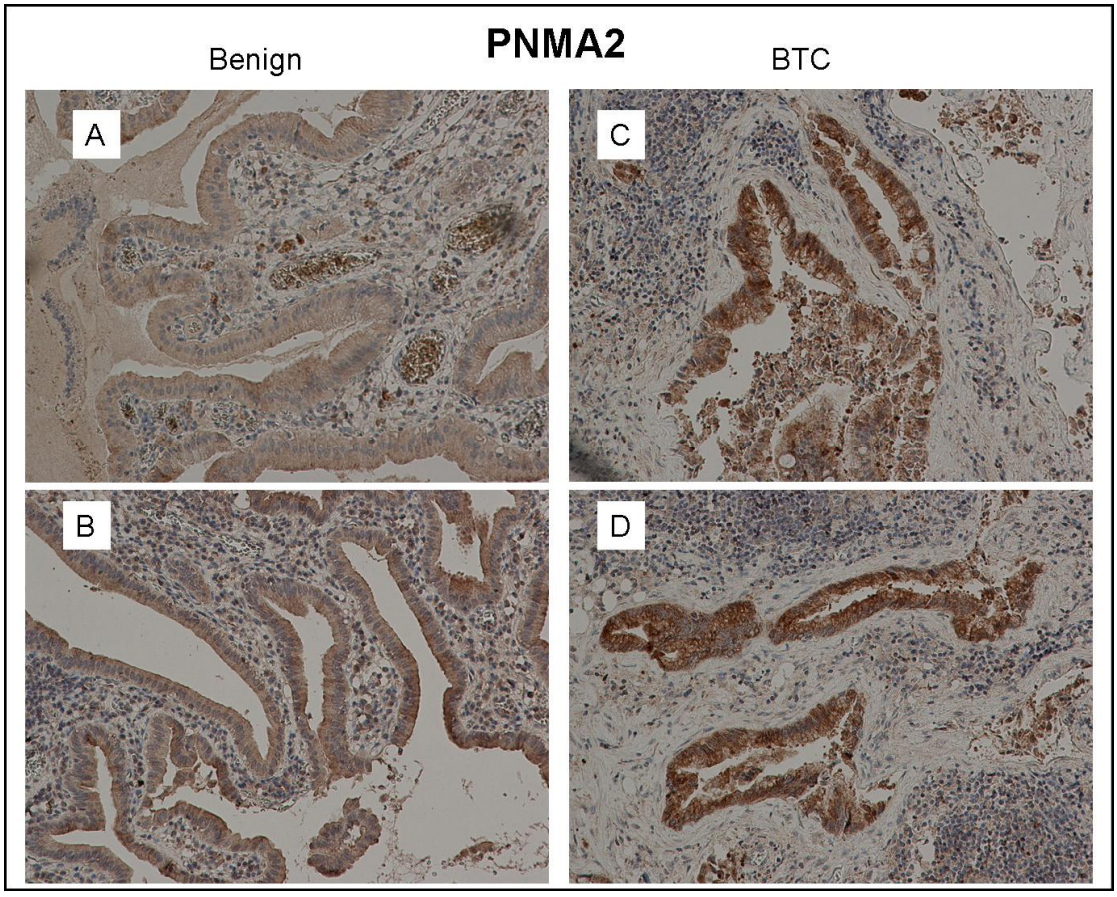


Figure 70. Examples of PNMA2 immunostaining showing negative or weak staining in benign disease (A&B) and moderate staining of epithelium in BTC (C&D).

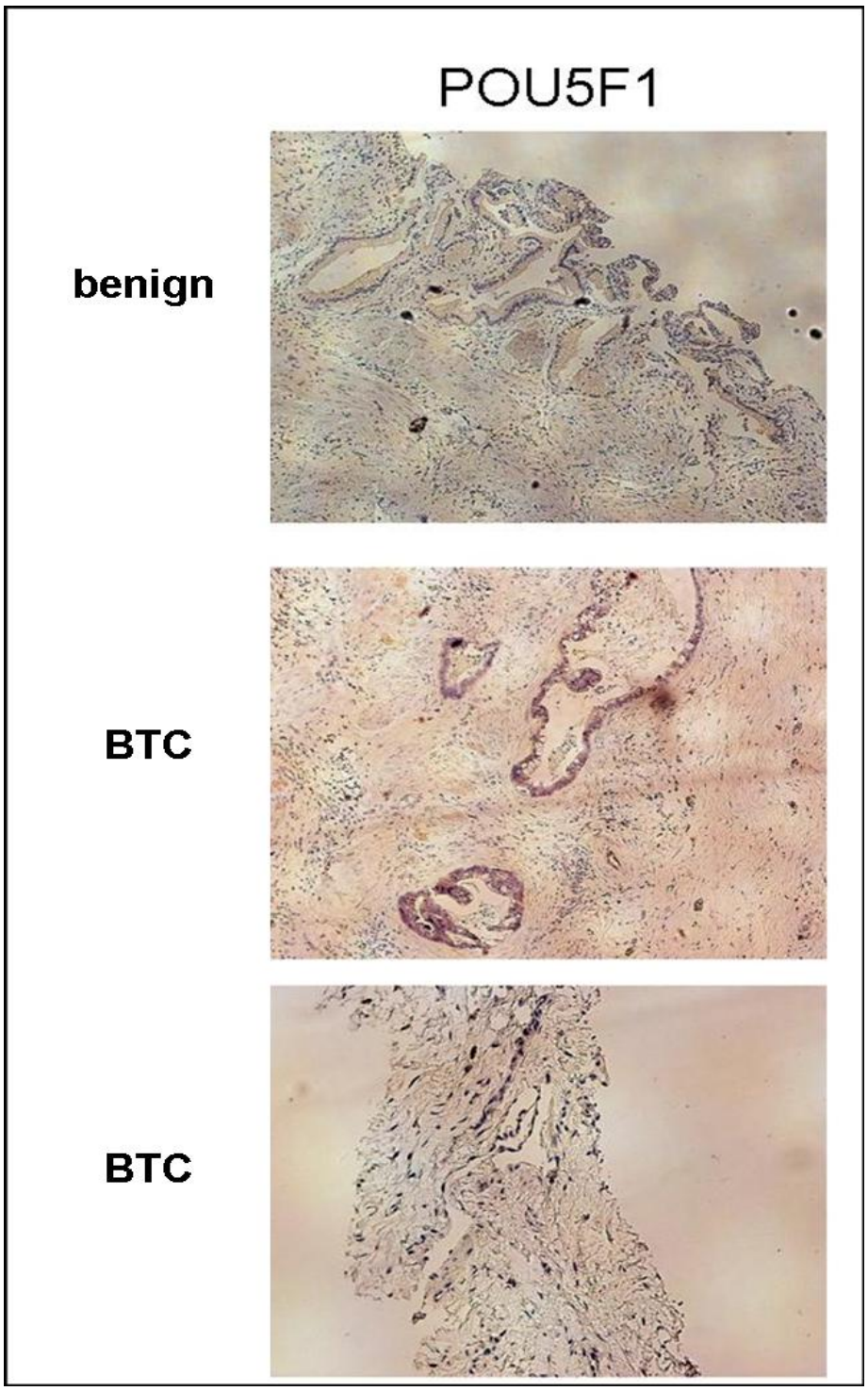


Figure 71. Examples of immunostaining for POU5F1 (oct-3/4). All benign and malignant samples were negative for immunostaining.

7.18 Conclusions: Analysis of protein expression by immunohistochemistry of biliary tissues.

The results of the immunohistochemistry support the hypothesis that upregulation of mRNA expression in genes of interest identified by biliary brushings +/- surgical resection material, is often translated to up-regulated protein expression in tissue sections. Out of the 4 proteins assessed to date, 3 were significantly up-regulated and the fourth showed no staining in any samples, raising the question of methodological problems with antigen exposure in the biliary tissue sections.

Although there were clear differences between the overall levels of staining and numbers of cells staining positive for CD9, SERPINA3 and PNMA2, none had sensitivity or specificity high enough to be diagnostic markers by immunohistochemistry. For example, using data for CD9, 10/12 cancer samples (83%) were positive for any staining and 2/11 (18%) were positive in the benign group, reducing its specificity. If one uses a stronger level of staining (≥2) as the cut-off, only 5/12 (42%) were positive in the cancer tissues and 1/11 (9%) were still positive in the benign tissues. Equally, of the cancer tissues, there was a wide range in the number of cells staining positive (5-70%), with up to 30% weak or moderate staining in the two benign samples staining positive for CD9. The data presented and numbers tested were insufficient to evaluate whether protein expression may be useful prognostically.

Despite the lack of high sensitivity or specificity by immunohistochemistry, the results support further evaluation of some of the proteins of interest in other clinical samples such as bile or blood using more quantifiable methodology such as ELISA.

7.19 Discussion: investigation of gene and protein expression in biliary surgical resection material.

These studies of mRNA expression in biliary tissues raise a number of interesting results. Analysis and confirmation of elevated MUC4, previously shown by our group to be upregulated in BTC, acted as an internal control supporting the validity of the results. Despite a strong correlation of mRNA expression in the second biliary brush validation set, the correlation was less strong in the biliary tissues, with a concordance of only 14%.

The lower concordance of upregulated mRNA expression in tissues compared to the ERCP biliary brushings may have a number of explanations. Firstly, we have shown that the relative abundance of epithelial and leukocyte RNA in the two types of tissues is different with the most striking difference being the higher quantity of leukocyte RNA in the tissue samples – in keeping with macroscopic assessment of biliary tissues where surgical resection material from patients with complex strictures and/or those who have biliary stents in situ, have much greater numbers of inflammatory cells than biliary brushings. These data suggest that brushings may be a purer source of epithelial RNA for experimental or clinical use.

When using ubiquitously expressed reference genes such as 18S or GAPDH, an alteration in the ratio of different sources of RNA may impact on the

relative expression of genes related to epithelial cell function or disease. One potential method of reducing this bias would be to use a standardised reference gene of epithelial cell origin. CK19 and some other cytokeratins are good biliary epithelial cell markers and are expressed at levels near to those of ‘house keeping’ reference genes such as GAPDH. Therefore, these could be used as reference genes for calculation of relative mRNA expression in the genes of interest. However, most of these genes are altered to some extent in disease states, including CK19 which has previously been shown to be upregulated in BTC (Maeda, Kajiyama et al. 1996), and would therefore also not be an ideal reference gene.

Another potential factor that we have not evaluated to date is the high stromal component and fibrotic nature of BTC tissues. These result in difficulties in breaking up the tissue during the initial steps of RNA isolation and may impair release of epithelial cells and the contained RNA. In addition, there may be a relatively high contribution of RNA from other stromal cellular components such as fibroblasts. These mesenchymal cells do not express CD45 used as our main marker of cells from haematogenous cell lineages. The number of fibroblastic cells is thought to be relatively low in comparison to epithelial cells and leukocytes but require further evaluation using fibroblast markers in biliary tissues.

One way of overcoming this problem would be to use laser capture microdissection of biliary tissues to isolate a purer pool of epithelial cells, although this introduces additional technical challenges.

An alternative explanation for the differences in gene expression between biliary brushings and tissues is that cells at the centre of the tumour mass (isolated from the tissue sections) have a different biology and gene expression to those isolated from brushings at the tumour surface and leading edge. This phenomenon of heterogeneity within cancers is well recognised and recently specifically investigated with regards different genetic clonal compositions within different areas of the same breast cancer mass from individual patients (Navin, Krasnitz et al. 2010). Also, biliary brushings isolate cells from the bile duct surface which are in contact with bile. It is highly likely that gene expression in the different areas of the tumour is different depending on the function of the genes of interest. This phenomenon may in fact also support the use of biliary brushings as a more suitable source of material for biomarker development as cells at the bile duct surface or leading edges may be more likely to secrete protein biomarkers into more practical clinical samples such as bile and serum.

Chapter 8.

8 Conclusions, discussion and future directions

8.1 Summary of main conclusions

1. Measurement of serum CK19 fragments (CYFRA 21-1) is shown to be a diagnostic and prognostic marker of BTC with accuracy similar to that of the current best biomarker, CA 19-9.
2. Measurement of serum CK19 fragments (CYFRA 21-1) as a surrogate marker of tumour cell fragments in the blood supports the concept that circulating tumour cells may be common in the blood of patients with BTC.
3. Although technically possible, mRNA expression profiling of archived samples of PSC related CC is currently not plausible, primarily due to the rarity of the samples and the very high (or complete) degradation of RNA caused by the methods used for storage. Alternative, RNA preserving methods of storage would optimise prospective sample collection for future studies.
4. RNA isolated from bile and biliary brushings obtained at ERCP is of low quantity and highly degraded but is suitable for gene expression analysis if appropriate techniques are used.

5. Whole genome mRNA expression profiling of biliary brushings is feasible and potentially provides a relatively simple and reliable method of assessing gene expression in biliary disease.
6. A selection of protein products of genes identified by the microarray and qPCR studies are shown to be upregulated in cancer tissues by immunohistochemistry, supporting the validity of the gene expression data.
7. Genes aberrantly expressed in BTC can be identified using brushings of biliary strictures. A selection of the genes identified in this thesis (e.g. PVT1, HOXA10, COL17A1) show promise as genes of interest in the biology or as biomarkers of BTC and warrant further investigation.

8.2 Discussion

Biliary tract cancer remains difficult to diagnose in many patients. Patients with BTC have a poor prognosis, in part related to difficulties and delays in diagnosis and most patients present at an advanced stage with five year survival rates of less than 5%. Diagnostic difficulties are related to the lack of reliable tumour markers, radiological similarities with benign disease such as PSC and IgG4 disease, low tissue diagnosis rates because of the low sensitivity of biliary brushings (18% to 69%, Table 1) and the absence of mass lesions that can be biopsied in many patients. Some surgical series report a high prevalence of benign disease (up to 17%) in patients undergoing resection for presumed CC, highlighting the need for accurate diagnosis in order to plan appropriate treatments (Erdogan, Kloek et al. 2008). High tissue diagnosis rates can be achieved (Abstract 2, Figure 74 (Mills, Chapman et al. 2009)) but there remains a need for improved biomarkers for BTC.

In Chapter 2, we demonstrate that a simple ELISA test measuring circulating fragments of CK19 (CYFRA 21-1) from epithelial cells, has a higher specificity (88% to 97%) but lower sensivity (30% to 56%) than serum CA 19-9, for the diagnosis of BTC. A major problem with all biomarkers is that until the specificity of the test reaches that of a cytologic or pathologic confirmation (ie usually 95% to 100%), it is difficult to recommend to patients high risk or potentially toxic treatments such as surgery or chemotherapy. Although the CYFRA 21-1 test has specificity approaching these figures, it is increasingly recognised that single biomarkers are unlikely to achieve such high specificities and be useful for diagnosis unless developed as part of multiple

biomarker tests that incorporate assays of different biological pathways or cellular components (Gervasoni, Monasterio Munoz et al. 2008). The diagnostic accuracy of CYFRA 21-1 can be improved (sensitivity 27% to 51%, specificity 95% to 100%) to levels similar to that of biliary brush cytology (sensitivity 18-69%, specificity 90-100%) by combination with CA 19-9, which recognises a different protein group in the form of extra-cellular mucins. The possibility of further improving diagnostic accuracy by addition of other biliary and/or epithelial markers remains but has not been tested.

The advantages of serum ELISA tests such as CYFRA 21-1, are that they are simple to perform, relatively cheap (approximately £10 per sample) and, unlike tissue diagnosis, are non invasive requiring only a sample of blood. Therefore, they have the potential to be effective and efficient tests for surveillance of high risk patients such as those with PSC. In order to address the use of CYFRA 21-1 in screening and diagnosis of BTC in patients with PSC, a large prospective multi-centre study is required which is feasible only as part of other collaborative studies of patients with PSC.

In Chapter 2, we also demonstrate that CYFRA 21-1 is a strong predictor of poor prognosis for patients with BTC (median survival 2 months v 10 months for levels ≥ & ≤ 3ng/ml respectively). These results are superior to that of CA 19-9, used by some as an indicator of advanced disease and non-resectability (Levy, Lymp et al. 2005). In our series, CA 19-9 was not a significant predictor of prognosis. The issue of whether CYFRA 21-1 in addition to staging by imaging modalities may be a marker of potentially curative surgery could not

be assessed with the small number of surgical cases in our samples store but deserves further consideration.

Another approach to biomarker development is to develop multiple gene signature assays using data pooled from gene expression studies for the disease of interest. Spira et al used this approach to identify microarray gene signatures in lung cancer (Spira, Beane et al. 2007). In this study, mRNA expression profiling of bronchial brushings from patients (all smokers) with and without lung cancer, could detect the presence of a cancer using an 80 gene signature with a sensitivity of 80% and specificity of 84%. Importantly, this disease identification was made by brushing macroscopically normal bronchial mucosa in patients with cancer elsewhere in the bronchial tree. Multiple gene assays are likely to be far more accurate for diagnosis of disease than measuring single gene or protein expression. As utilised and demonstrated in Chapter 6, assays are now available that can reliably measure multiple gene expression using small amounts of mRNA. Such assays can easily be customised to assess genes and/or splice variants of interest and are becoming simpler and cheaper to use. They therefore lend themselves to further evaluation as diagnostic or prognostic gene expression assays.

An initially unexpected finding of our work was the level of RNA degradation found in the biliary brush samples. However, using methodology developed for use with similarly degraded RNA from FFPE tissues, we were able to show that reliable data can be extracted from samples of biliary brushings. This, to

our knowledge, has been little studied in the past and is a particular strength of this thesis. We also showed that, in contrast to earlier studies reporting that both bile and x-ray contrast agents are cytotoxic to biliary epithelial cells (Benedetti, Alvaro et al. 1997; Ju, Kim et al. 2002), the degradation does not appear to be related to short term exposure of cells to bile or x-ray contrast. Also the degradation occurs *in vivo* and is thus not amenable to methodological changes to prevent it.

We have shown that whole genome mRNA expression profiling using microarray, supported by qPCR, can be done using biliary brushings. Moreover, our data on relative expression of epithelial (CK19) and leukocyte (CD45) markers in biliary brushings and surgical resection tissue suggest that biliary brushings are a more reliable source of epithelial cells than tissue. This is in line with microscopic findings at cytologic or histologic assessment of biliary brushings and resection tissues which tend to have a greater number of inflammatory cell infiltrates in the complex strictures of patients undergoing surgery. Also, samples of biliary brushings are more clinically relevant samples for biomarker work as ERCP is amongst the earliest tests done in most patients with indeterminate strictures and is less hazardous than percutaneous biopsy or surgery.

Our work is one of the first thorough evaluations of the suitability of biliary brushings for gene expression studies. As such, data presented in this thesis support the use of these methodologies for further, large-scale studies in patients with indeterminate biliary strictures. Also, these methods could be

applied to other biliary diseases that have historically been very difficult to study, such as PSC, PBC and IgG4 disease. One of the major problems in the past has been the lack of biliary epithelial cells for study as they are not amenable to sampling by biopsy and are difficult to isolate by laser capture microdissection (LCM) in the minority of cases that undergo surgery. To date, we are aware of only a few studies where specific analysis of biliary epithelial cells has been undertaken following LCM of surgical resection tissues in PSC and PBC (Taniai, Higuchi et al. 2002; Baba, Kobashi et al. 2006). However, with the development of more direct and sophisticated ERCP techniques, including cholangioscopy, it is likely that directed biopsies of intraluminal pathologies would provide a similar or even better source of material for work similar to that described in this thesis.

PSC can be a difficult condition to manage and patients have a high risk of developing cholangiocarcinoma. This is a topic of particular interest to us and developing an understanding of the biology and biomarkers for use in this clinical setting were initially major aims of these studies. We investigated the feasibility of using archived FFPE tissues from patients with PSC and PSC-related CC to achieve these aims. However, as described in Chapter 3, despite methodology now being available to perform such studies, the clinical materials were too few and were unsuitable for further analysis using gene expression studies.

The phenomenon of a dysplastic ‘field change’ and the hypothesis of precancerous dysplastic change in conditions such as PSC (Fleming, Boberg

et al. 2001), led to our decision to transfer work to the biliary tree and investigate gene expression in biliary brushings which commonly have no malignant cells, even when directly brushing malignant strictures. Recent data assessing explants of 100 PSC patients who underwent liver transplantation reported that dysplasia is seen in bile ducts of 83% of those with coexistent CC and 36% of those without CC (Lewis, Talwalkar et al. 2010). This high prevalence remained when recording only high grade dysplasia which was seen in 72% and 32% in those with and without co-existent CC in the explant. Assessing gene expression and DNA mutational analysis in biliary samples from patients with PSC is difficult because of the complexities of isolating biliary tissues. One approach is to use laser capture microdisseaction of liver transplant explants, but these tissues are not common and represent advanced disease. Methods described in this thesis provide a technique for sampling and evaluation of this problem in patients with PSC before or during evaluation for liver transplantation. We support such studies in the form of prospective sample collection of biliary brushings from patients with PSC and other biliary diseases, who undergo ERCP.

Cholangitis in patients with biliary strictures or stents remains a potential problem for gene expression profiling of biliary samples. A potential confounding effect of infection in patients with complex malignant strictures cannot be excluded but we demonstrate two important points that suggest that any such confounding effect is small. The first is that the relative quantities of epithelial and leukocyte RNA is such that differential leukocyte expression between groups is likely to be minimised by the abundance of

epithelial RNA which probably accounts for approximately 50 times more RNA than leukocyte RNA. Secondly, the microarray data failed to show differential expression of genes commonly used as leukocyte specific markers such as CD45, leukocyte specific protein-1 (LSP1), leukosialin (CD43), CD18 (MHM23), cathepsin G, leukocyte alkaline phosphatase (LAP), CD11 and CD166. We are therefore confident that the potential confounding effect from leukocytes remains small.

Another potential source of error in our data is the relatively small sample sizes. This is particularly true for the large data sets obtained from the microarray experiments. However, such samples are precious and costs for microarray analyses are high. Therefore, feasibility studies, such as these presented in this thesis must be performed before large prospective studies could be planned. Following bioinformatics advice, appropriate statistical methodologies were applied to minimise bias and reporting errors, although these cannot be removed completely with the small sample sizes used in our microarray work. The validity of the results with regards sample size errors was also supported by the high concordance between the biliary brush validation sets as well as the finding of up-regulated protein products by immunohistochemistry of tissue sections. This work continues with assessment of HOXA10 and COL17A1, both of which were found to be highly upregulated in all of the sample sets.

The lower concordance between the biliary brush and tissue gene expression data may at first appear disappointing. However, as discussed, these two

clinical samples are quite different and one could not expect the expression pattern to be exactly the same. The higher ratio of leukocyte and stromal material in tissue sections is clearly a problem in these samples. Also, the biology of cells in the centre of masses (used for the tissue work) is somewhat different to that of the cells lining the bile duct isolated by biliary brushings. Both have advantages and disadvantages. The surface epithelial cells in bile ducts are those most likely to be relevant for investigation of potential protein products released into the bile which may be a good source of material for biomarker work using ELISA, but, as discussed in Chapter 7, less suitable for Western blot.

The relative quantities of fibroblasts in benign and malignant samples have not been assessed in our work. However, based on cytologic and histologic assessment, this is likely to be small in samples of biliary brushings but may be more significant in tissues sections of BTC, which are known to have relatively high proportions of stromal material (Okamura, Yoshida et al. 2005). Fibroblastic and stromal tissue activity may be important in both understanding the biology of cancer and in biomarker development. Studies in other malignancies, such as breast and colon cancer, first demonstrated alterations in stromal cell numbers and functions in cancer biology (Allinen, Beroukhim et al. 2004; Nakagawa, Liyanarachchi et al. 2004). More recent work in pancreatic and biliary tract cancer suggest that fibroblast activity and gene expression becomes altered and may play a role in disease progression (Hwang, Moore et al. 2008; Chuaysri, Thuwajit et al. 2009; Utispan, Thuwajit et al. 2010).

Another potential validation step, which may also be a possible clinical diagnostic test, is for immunocytochemistry to assess protein expression in the biliary brushings, samples which closely match the samples used for the gene expression studies. A major obstacle for this however is the paucity of the material obtained from biliary brushings, which is usually sufficient for smearing only 2 to 4 slides, and hence insufficient for research work without additional sampling of strictures with a second or third brush of bile ducts.

In summary, the data presented in this thesis describe the methodology for gene expression studies in biliary brushings, data on gene expression in BTC, and evidence to support this methodology and the validity of the results. Such methods should be suitable for investigation of other biliary diseases.

8.3 Plans for further experiments

Further experiments are required in order to investigate the biology of genes of interest but in particular to test the hypothesis that some of these genes may serve as useful biomarkers in BTC. These were not completed as part of this thesis but represent areas of planned and ongoing research by our group.

8.3.1 BTC cell lines

Some of the genes identified thus far have not been previously reported as biomarkers or investigated with regards the biology of BTC and/or other cancers. A better understanding of the biology of BTC and cancers in general

is likely to lead to the development of new cancer therapies and such work firstly requires investigation *in vitro* using suitable cell lines such as the TFK-1 and HUCCT BTC cell lines available in our group.

From the shortlist of over-expressed genes further assessed in this thesis, PVT1 is particularly interesting with regards the biology of BTC. PVT1 (also known as MYC activator) is a non protein-coding gene that was initially identified as associated with lymphoma where chromosomal translocations result in PVT1 becoming sited near to and co-amplified with the MYC oncogene. It is thought to play a role in the normal MYC pathway via microRNAs encoded within it (Beck-Engeser, Lum et al. 2008). Its location on chromosome 8q24, a locus commonly amplified in other solid organ cancers, led to the investigation of PVT1 in breast and ovarian cancer cell lines where it was shown that RNA silencing of PVT1 results in apoptosis of cancer cell lines, a phenomenon that was not seen with silencing of MYC (Guan, Kuo et al. 2007). BTC is a relatively chemotherapy resistant cancer and elucidation of mechanisms responsible for sensitising cells to chemotherapy +/- other treatments such as PDT, may result in the generation of new therapies in BTC. Using techniques and conditions already set up by the group, we aim to assess the effect of RNA silencing PVT1 in BTC cell lines and assess the response to addition of other treatments such as EGFR blockade and PDT.

Similarly, HOXA10 may be important in the biology of BTC. The HOX genes are particularly important in embryogenesis but become less active in normal adult physiology. HOXA10 is strongly expressed in a number of tissues during

development but is mostly only weakly expressed in normal adult tissues apart from endometrial cells in the proliferative mid-secretory phase of the menstrual cycle. The primary drive for activation in normal physiology is thought to be related to female sex hormones. HOX genes are not expressed in normal ovarian surface epithelium but some (including HOXA10) are expressed in some ovarian cancers (Cheng, Liu et al. 2005). In haemopoietic cell lines, induction of HOXA10 over-expression leads to increased proliferation of haemopoietic cell lines and transformation to cells with stem cell like properties, including self renewal (Magnusson, Brun et al. 2007). A recent report also demonstrates up-regulation of HOXA10 as a marker of oral squamous cell cancer (Yamatoji, Kasamatsu et al. 2010).

We aim to assess the expression of HOXA10 (and possibly other HOX genes such as HOXB6 shown to be upregulated in BTC by our and other data) in BTC cell lines. If shown to be up-regulated, we would again use RNA silencing to assess the effect on cell survival, with or without the addition of other therapies such as chemotherapy, EGFR blockade and PDT.

8.3.2 Measurement of COL17A1 or COL17A1 fragments in bile and serum

Collagen 17α1 (COL17A1) is an atypical collagen in that it is a membrane bound protein rather than a component of the extra-cellular matrix. It forms part of the hemidesmosomes and is involved in cell adhesion to the basement membrane. It has a secreted,180kD, extra-cellular domain (also known as

BP180, ectodomain or LAD-1) releasing soluble collagen fragments into surrounding tissues. To date, reports of abnormalities of COL17A1 function are limited to the benign skin condition epidermolysis bullosa and there are no published reports of altered COL17A1 in cancers.

Our data show that COL17A1 is highly and consistently up-regulated in all the sample sets tested (microarray x16, qPCR validation of brushings x11, qPCR with fresh set brushings x5, qPCR tissue x16). These data are in line with a small study (n=17) demonstrating higher levels of type IV collagen in embryologically related pancreatic adenocarcinoma (Ohlund, Lundin et al. 2009). This study used an ELISA assay to measure plasma type IV collagen and showed significantly higher levels in cancer and a strong association between elevated levels and poor prognosis. Similar results for another collagen (collagen XXIII) have been reported in prostate cancer where elevated levels were associated with both diagnosis and poor prognosis (Banyard, Bao et al. 2007).

The striking pattern of mRNA up-regulation and the potential that the extra-cellular domains may be secreted into bile or serum, support further investigation of the hypothesis that COL17A1 may be a biomarker for BTC. We aim to study this possibility by assessment of protein levels in tissue using immunohistochemistry, and in bile and serum using antibodies directed against the extra-cellular domain of the protein by Western blot and ELISA assays. Commercial kits are currently not available but antibodies directed against the extra-cellular portion have been described (Schacke et al 1998).

Thus, it should be possible to develop an ELISA assay in order to measure protein levels and test as a diagnostic and/or prognostic marker. Similarly, it may be useful to investigate the protein expression of other collagens (COL1A1 and COL6A3) also shown by our data to be consistently up-regulated in BTC.

8.3.3 Analysis of further genes identified in the microarray data set of BTC biliary brushings.

The original data set obtained from the microarray analysis of the biliary brushings contained a list of over 2000 genes reported to be upregulated in BTC. Many of these have only a small level of fold change or statistical significance but a number of genes remain of interest for further investigation in BTC. Plans are in place to investigate some of these genes using the methods outlined in this project. Examples include MLL4 (myeloid-lymphoid leukaemia 4, also known as TRX2, MLL2) which was originally identified as amplified in leukaemia but subsequently shown to be a commonly amplified locus (19q13.1) in solid organ tumours including some pancreatic cancers (Huntsman, Chin et al. 1999). Our data uaing biliary brushings suggest that MLL4 is upregulated in BTC (fold change 7.7, $p<0.001$) and data in proteinatlas.org suggest weak or negative staining in normal bile ducts but positive staining (>75%, moderate) in CC. Based on our microarray data, biological function and/or proteinatlas.org data, other genes deserving further investigation include THBS1, TSP1, PPFIBP1, RAPGEF3, ADAMTSL4, CAMK1G, IFITM2, MED12 and NGFRAP1.

8.3.4 Identification and analysis of circulating tumour cells in BTC

CYFRA 21-1 is a component of the intracellular cytoskeletal protein CK19, present in all epithelial cells. Due to the fact that epithelial cells ordinarily are not present in the blood of healthy individuals or those wth benign diseases, CK19 and/or CYFRA 21-1 may act as surrogate markers for cancer cells in blood. Our data assessing the role of CYFRA21-1 in BTC demonstrate that it is a relatively good biomarker for BTC (Chapter 2). Sensitivity and specificity were 30% to 56% and 88% to 97%, respectively – figures similar to those for the biomarker CA19-9, which is currently in widespread clinical use. However, these data also support the hypothesis that there may be significant numbers of circulating tumour cells in the blood of patients with BTC, a phenomenon that has been little investigated and deserves further investigation.

One explanation for the high prevalence of circulating CYFRA 21-1 may be that our cohort included mostly patients with advanced disease. However, similar findings were reported in a pre-operative surgical cohort of Japanese patients with early stage disease (Uenishi, Yamazaki et al. 2008). Another finding of both our work and that of Uenishi et al was that CYFRA 21-1 is a strong predictor of poor prognosis, and may play a role in the pre-operative assessment of resectability. This may be related to CYFRA 21-1 acting as a marker for the presence of circulating tumour cells in those with a poorer outcome.

These early findings support the plan for further experiments to assess circulating tumour cells in BTC. Sensitive, automated technology is now available that can isolate CTCs from peripheral blood and we plan to use these assays in a patient cohort with benign and malignant biliary disease. The current best example of CTC capture technology is the CellSearch assay (Veridex LLC, Raritan, NJ, USA) which has now been in use for over 5 years. The assay relies on magnetic particles coated with a selection of epithelial (EpCAM, CK8, CK18 & Ck19) and leukocyte (CD45) markers that can be used to separate cells from whole blood. Reports consistently show high recovery rates of cancer cells spiked into healthy blood and negative or very low rates (<1% of samples) of circulating epithelial cells in healthy or benign diseases. For example, a large early study of 964 patients with different cancers and 344 healthy or benign disease controls demonstrated a very high recovery rate (85%-122%) of cancer cells (range 4-1200 cells) spiked into 7.5mls of healthy blood, with no CTCs found in healthy controls and only 1 low level of CTC found in 1 of 199 disease controls (Allard, Matera et al. 2004). The same study reported CTCs in 36% of all the 964 assorted adenocarcinomas which also included identifiable CTCs in 4 (19%) of 16 patients with pancreatic adenocarcinoma, a cancer with embryologic and morphologic similarities to BTC. Some cancers such as breast and prostate cancer, appear to more commonly have CTCs in the blood (eg 61% in breast cancer (Cristofanilli, Budd et al. 2004), 57% in prostate cancer (Allard, Matera et al. 2004)) than other cancers. There are few data in pancreatic cancer suggesting CTCs being identified in 19% to 48% of cases (Allard, Matera et

al. 2004; Ko, Scott et al. 2007). However, we are aware of no reported data for analysis of CTCs in BTC to date.

Obvious clinical applications of these technologies are to further investigate measurement of CTCs (along with serum CYFRA 21-1) as markers of diagnosis and prognosis in patients with BTC as well as investigation of its role in determining resectability and surgical outcome. Protocols for these studies are being developed in surgical and palliative patients with BTC as well as pancreatic adenocarcinoma and could be extended to patients with PSC-related or other indeterminate biliary strictures.

Another potential use of these assays is to isolate CTCs for further investigation of the biology of BTC and for biomarker work. Gene expression analysis can be performed using the small numbers of cells isolated with the use of modern, sensitive RNA isolation and amplification kits. Several studies have now been published demonstrating that CTCs can be used for multiple gene expression analysis by qPCR (O'Hara, Moreno et al. 2004) and by whole genome microarray analysis (Smirnov, Zweitzig et al. 2005). In addition to the general investigation of BTC biology, there is the potential for assaying individual patient samples to assess whether they may be suitable for targeted individualised therapies using particular chemotherapy and biologic regimens (Perou, Sorlie et al. 2000; Dressman, Hans et al. 2006).

8.3.5 Detection of circulating mRNA of epithelial cell markers as biomarkers for BTC

Following a similar hypothesis to the measurement of circulating epithelial cell protein fragments such as CYFRA 21-1, an alternative approach to identifying intact CTCs is to measure circulating levels of epithelial and cancer related mRNA. This is now feasible using sample collection and RNA preservation techniques and has been successfully demonstrated for other cancers. In a study of 68 patients with pancreatic cancer and 31 healthy or benign disease controls, CK19 mRNA was elevated in 64% of blood samples from patients with pancreatic cancer when compared to the maximum level identified in the benign controls (Hoffmann, Kerner et al. 2007). However, it is unclear currently what represents 'normal' levels as demonstrated by another study where CK20 mRNA in blood was reported in 34% of 154 patients with pancreatic cancer but also reported as present in 17% of 54 benign disease (chronic pancreatitis) controls (Soeth, Grigoleit et al. 2005). A recent study addressed the question of comparison between different methods for the detection of CTCs demonstrating a reasonable correlation (72% agreement, K coefficient 0.356) using the CellSearch assay and qPCR for CK19 mRNA in blood (Van der Auwera, Peeters et al. 2010).

High throughput qPCR techniques using 96- or 384 well plates and well tested CK19 qPCR primer sets allows relatively easy measurement of circulating CK19 mRNA in clinical samples in our laboratory. Measurement of circulating CK19 mRNA in the blood of patients with benign and maligment biliary

diseases may be possible using blood samples already in storage. However, the rapid rate of RNA degradation by RNase activity and other means in blood will likely result in low levels of positive results. This could be easily tested with pilot studies, possibly using the same samples as the CYFRA 21-1 study as a direct comparison. Low levels of CK19 mRNA amplification may represent low levels of CTCs in BTC or degradation of the mRNA requiring the collection of a fresh sample set with the addition of RNase inhibitors in the collection tubes.

8.3.6 Validation of protein biomarkers using clinical samples of blood and bile

The final aim of this project was to validate the clinical effectiveness of the most promising biomarkers in clinical samples stored in our tissue bank. We will assess a selection of the biomarkers for their diagnostic and prognostic accuracy in patients with BTC. Studies will begin with the measurement of COL17A1 levels in bile and/or blood. Other potential protein markers may follow later. Another important clinical problem is appropriate surveillance of CC in patients with PSC. We aim to test the effectiveness of the biomarker(s) in patients with PSC and PSC related CC. A further level of validation may require larger sample numbers in the form of multicentre studies which would be feasible with collaborative work on sample collection in patients with BTC and PSC.

8.4 Other potential future directions

8.4.1 Multicentre studies of biomarkers in patients with BTC and PSC

Most studies on biomarkers in BTC involve relatively small numbers with a large range in reported accuracy of such markers. The only way of gaining sufficient numbers for adequately powered studies and to obtain clear results is to undertake prospective, multicentre studies involving sample collection and storage. Such studies are feasible and are planned for patients with BTC and those with PSC.

8.4.2 Gene expression profiling of PSC related biliary strictures

The biology of PSC is poorly understood for a number of reasons including the relative rarity of biopsy material, difficulty in isolation and culture of normal or PSC related cholangiocytes and the small volume of the affected tissue for sampling (i.e. biliary epithelium). The collection and processing of biliary samples described in this project provides a method for further investigation of disease biology using techniques such as gene expression profiling and DNA mutational analysis.

8.4.3 Gene expression profiling of radiographically or cholangioscopically directed intra-ductal biliary biopsies.

Some centres, including our centre at UCH, use radiographically directed intra-ductal biopsies taken from strictures at the time of ERCP. Our own data (see Abstract 2 (Figure 74) in appendix) and those of others suggest that the sensitivity for confirming the diagnosis of BTC is between 44% and 53%, higher than achieved in most studies of biliary brush cytology (Weber, von Weyhern et al. 2008; Mills, Chapman et al. 2009). The recent introduction of cholangioscopy has led to the possibility that similar work could be carried out using directed biopsies of malignant biliary strictures. This may provide more reliable sample acquisition. However, as yet, there are insufficient data to demonstrate that cholangioscopically directed biopsies provide more accurate tissue sampling and diagnosis. Also, the data on a field change in macroscopically normal tissue taken at the time of bronchial brushings in patients with lung cancer, and from patients undergoing surgery for hepatocellular carcinoma suggest that directed biopsies may not be required.

8.4.4 Diagnostic gene chips for BTC

Most diagnostic tests to date involve measurement of protein levels. However, methods such as qPCR are being simplified and standardised so that PCR based tests are now available, for example in identification of bacterial DNA or quantification of viral load in hepatitis B and C infection. Work using microarray gene expression profiling has been applied in the clinical setting in

patients with breast cancer in order to stratify patients at risk of metastases and to direct decisions regarding the use of adjuvant chemotherapy (Perou, Sorlie et al. 2000; Dressman, Hans et al. 2006). Similar work could be applied to patients with hepatobiliary malignancy and may also include assessment of protein receptor activity for tailoring therapies such as addition of biological agents.

Now that techniques for RNA isolation and qPCR are also being simplified with kits and chips, it has become realistic to perform such work using clinical samples for diagnostic purposes. An obvious extension of our work is to use whole genome expression data to design diagnostic qPCR chips or protein arrays customised for a selection of sensitive and specific markers. The accuracy of such chips is likely to be greatly improved by the addition of multiple (24 to 96) genes or proteins and could reach levels that may mean that biopsy confirmation is not required. Also, it is possible that disease in high risk groups such as those with PSC, may be identified at a premalignant or early malignant phase where curative treatments such as surgery or transplantation may be appropriate.

8.4.5 miRNA profiling

Micro RNA (miRNA) are short RNA species that alter mRNA activity at the post translational stage. The number of miRNA is far smaller than the number of genes or mRNA but each miRNA interacts with many different mRNA molecules with wide biological effects. miRNA have been implicated in many cancers and have also been shown to be useful biomarkers for some cancers

including cholangiocarcinoma (Meng, Henson et al. 2006). They also provide a potential therapeutic approach as activation or inhibition of some miRNA has been shown to alter cancer cell biology *in vivo* (Meng 2006).

Recent investigations have shown that microRNAs are now measurable in blood and have the potential for being effective biomarkers for diagnosis, prognosis and screening of patients with cancer. Examples include studies in pancreatic cancer (Jones, Zhang et al. 2008) and a study investigating a panel of 95 microRNAs in 90 patients with colorectal cancer in which a single microRNA (miR-92) could distinguish patients with colorectal cancer from those with inflammatory bowel disease (n=20) and healthy controls (n=50) (sensitivity 89%, specificity 70%, AUC 88.5%) (Ng, Chong et al. 2009). During this project we considered the option of miRNA profiling in biliary brushings, but the methods we used for mRNA isolation unfortunately result in the loss of very small RNA species (less than approximately 50bp), including miRNA, which require different isolation techniques. However, the small size and nature of miRNA mean that they are relatively resistant to degradation and therefore may be a very useful material to investigate in bile or biliary brushings. Microarray and qPCR arrays for investigating miRNA expression are available and could be used in a similar way to that outlined in this project.

Another important finding in miRNA research is that their relative resistance to degradation means that miRNA can be identified in relatively high quantities in peripheral blood. (Mitchell, Parkin et al. 2008). This provides an opportunity

for using easily accessible samples of blood for miRNA profiling in patients with benign and malignant biliary disease

8.4.6 DNA mutation and methylation status analysis

The biliary brush cytology samples in TRI Reagent had the RNA recovered but the DNA remains stored in a -80^{0}C freezer. The DNA could be easily separated from the TRI Reagent and used for further applications such as analysis of DNA mutations or methylation status in BTC.

8.5 Final concluding remarks

In summary, we present data detailing the mRNA expression profiling of ERCP biliary brushings in BTC with further validation of a selection of identified genes at the protein level. We have shown that the the degraded nature of the RNA isolated from biliary brushings that previously hampered accurate gene expression profiling, can be overcome using current techniques of molecular biology, with the most important factor being the use of short nucleotide sequences for measurement of gene expression. We demonstrate that samples of biliary brushings may be a better source of material than archived FFPE tissue sections for investigation of biliary disease. The results provide provisional data on potential novel biomarkers in BTC including COL17A1, CD9 and HOX genes. These require further larger validation studies before they can be studied in clinical practice. As far as we are aware, this is one of the first comprehensive studies of gene expression of ERCP biliary brushings and the methods described in this thesis may provide a platform for further investigation of BTC and other biliary diseases such as PSC and IgG4 disease in the future.

9 Appendix

9.1.1 Raw qPCR threshold data from the TaqMan Array experiments and calculations of relative expression

Table 31. TaqMan Array qPCR calculations of microarray set biliary brush samples

		ΔC_T (GAPDH-18S)		ΔC_T (GAPDH-18S)		$\Delta\Delta C_T$ $\Delta C_T - \Delta C_{T,cal}$		GAPDH rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	15.07	0.12							
	DA	13.69	0.19							
	JP	14.45	0.08				0.07	1.00	1.05	0.95
	PK	13.23	0.18	14.11	0.07	0.00	0.10	1.00	0.07	0.05
BTC, n=5	JH	14.08	0.25							
	LB	13.60	0.04							
	MT	15.22	0.30							
	PS	14.70	0.15				0.09	1.00	1.06	0.94
	RB	12.92	0.06	14.11	0.09	0.00	0.11	1.00	0.08	0.06

		ΔC_T (18S-18S)		ΔC_T (18S-18S)		$\Delta\Delta C_T$ $\Delta C_T - \Delta C_{T,cal}$		18S rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1									
	DA									
	JP						#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
	PK			#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
BTC, n=5	JH									
	LB									
	MT									
	PS						0.09	#DIV/0!	#DIV/0!	#DIV/0!
	RB			#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!

		ΔC_T (MUC4-18S)		ΔC_T (MUC4-18S)		$\Delta\Delta C_T$ $\Delta C_T - \Delta C_{T,cal}$		MUC4 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	26.03								
	JP	20.88	0.45				0.35	1.00	1.27	0.79
	PK	21.85	0.53	23.16	0.35	0.00	0.49	1.00	0.34	0.24
BTC, n=5	JH	19.95	0.09							
	LB	16.11	0.05							
	MT	17.83	0.53							
	PS	18.44	0.15				0.09	29.37	31.18	27.66

	RB	19.10	0.15	18.29	0.12	-4.88	0.37	29.37	7.47	2.34
						$\Delta\Delta C_T$				
		ΔC_T (MUC5AC-18S)		**ΔC_T (MUC5AC-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**MUC5AC rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	18.50	1.40							
	DA	18.05	0.08							
	JP	13.84	0.11				0.36	1.00	1.28	0.78
	PK	13.60	0.24	16.00	0.36	0.00	0.50	1.00	0.35	0.25
BTC, n=5	JH	14.40	0.09							
	LB	12.33	0.12							
	MT	10.38	0.18							
	PS	15.94	0.22				0.06	7.41	7.75	7.09
	RB	12.49	0.07	13.11	0.06	-2.89	0.36	7.41	1.86	0.33
						$\Delta\Delta C_T$				
		ΔC_T (ACTB-18S)		**ΔC_T (ACTB-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**ACTB rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	12.02	0.12							
	DA	13.03	0.07							
	JP	12.93	0.08				0.04	1.00	1.03	0.97
	PK	11.40	0.08	12.34	0.04	0.00	0.06	1.00	0.04	0.03
BTC, n=5	JH	11.25	0.06							
	LB	10.29	0.20							
	MT	12.49	0.13							
	PS	12.55	0.15				0.06	2.03	2.11	1.95
	RB	10.03	0.02	11.32	0.06	-1.02	0.07	2.03	0.10	0.08
						$\Delta\Delta C_T$				
		ΔC_T (CD9-18S)		**ΔC_T (CD9-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**CD9 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	18.21	0.38							
	DA	18.00	0.07							
	JP	17.05	0.08				0.11	1.00	1.08	0.93
	PK	17.07	0.16	17.58	0.11	0.00	0.15	1.00	0.11	0.07
BTC, n=5	JH	16.98	0.12							
	LB	14.99	0.06							
	MT	16.86	0.15							
	PS	18.09	0.16				0.05	2.15	2.23	2.07
	RB	15.48	0.07	16.48	0.05	-1.10	0.12	2.15	0.18	0.08
						$\Delta\Delta C_T$				
		ΔC_T (Notch3-18S)		**ΔC_T (Notch3-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**Notch3 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	23.79	0.26							
	JP	21.45	0.37				0.18	1.00	1.13	0.88
	PK	21.05	0.30	22.55	0.18	0.00	0.25	1.00	0.18	0.12
BTC, n=5	JH	21.02	0.14							
	LB	17.36	0.13							
	MT	20.68	0.32							
	PS	21.41	0.19				0.16	8.01	8.96	7.16

	RB	17.27	0.69	19.55	0.16	-3.00	0.24	8.01	1.34	0.90
						$\Delta\Delta C_T$				
		ΔC_T (ASPHD1-18S)		**ΔC_T (ASPHD1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**ASPDH1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.88	0.98							
	DA	18.32	0.10							
	JP	20.47	0.18				0.25	1.00	1.19	0.84
	PK	17.58	0.07	19.06	0.25	0.00	0.35	1.00	0.25	0.17
BTC, n=5	JH	19.21	0.09							
	LB	17.02	0.33							
	MT	18.73	0.14							
	PS	20.56	0.18				0.08	1.65	1.75	1.56
	RB	16.18	0.02	18.34	0.08	-0.72	0.26	1.65	0.30	0.09
						$\Delta\Delta C_T$				
		ΔC_T (ATP6V0A2-18S)		**ΔC_T (ATP6VOA2-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**ATP6V0A2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.56	0.29							
	DA	19.97	0.10							
	JP	21.12	0.36				0.12	1.00	1.09	0.92
	PK	19.24	0.15	19.97	0.12	0.00	0.17	1.00	0.12	0.08
BTC, n=5	JH	19.75	0.08							
	LB	18.71	0.17							
	MT	20.50	0.40							
	PS	21.38	0.23				0.11	1.10	1.18	1.02
	RB	18.84	0.17	19.84	0.11	-0.14	0.16	1.10	0.12	0.08
		ΔC_T (CEACAM1-18S)		**ΔC_T (CEACAM1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**CEACAM1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	18.25	0.05							
	DA	20.31	0.06							
	JP	20.86	0.08				0.09	1.00	1.06	0.94
	PK	20.02	0.32	19.86	0.09	0.00	0.12	1.00	0.08	0.06
BTC, n=5	JH	20.60	0.11							
	LB	18.28	0.04							
	MT	19.42	0.13							
	PS	19.05	0.17				0.05	2.08	2.16	2.00
	RB	16.67	0.12	18.80	0.05	-1.05	0.10	2.08	0.15	0.08
		ΔC_T (CELSR1-18S)		**ΔC_T (CELSR1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**CELSR1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	24.28	0.27							
	JP	22.94	0.08				0.11	1.00	1.08	0.93
	PK	21.56	0.18	23.17	0.11	0.00	0.16	1.00	0.11	0.08
BTC, n=5	JH	20.85	0.58							
	LB	19.37	0.11							
	MT	19.72	0.40							
	PS	23.09	0.17				0.15	7.15	7.92	6.45
	RB	18.64	0.08	20.33	0.15	-2.84	0.19	7.15	0.92	0.73

		ΔC_T (CFDP1 - 18S)		ΔC_T (CFDP1-18S)		ΔC_T - $\Delta C_{T,cal}$		CFDP1 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.05	0.79							
	DA	19.02	0.07							
	JP	20.33	0.11				0.21	1.00	1.15	0.87
	PK	17.14	0.24	18.88	0.21	0.00	0.29	1.00	0.20	0.14
BTC, n=5	JH	19.16	0.21							
	LB	17.12	0.06							
	MT	19.40	0.13							
	PS	20.19	0.20				0.07	1.11	1.16	1.06
	RB	17.78	0.03	18.73	0.07	-0.15	0.22	1.11	0.17	0.05
		ΔC_T (COL17A1-18S)		ΔC_T (COL17A1-18S)		ΔC_T - $\Delta C_{T,cal}$		COL17A1 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.88								
	DA	25.32	0.27							
	JP	21.28	0.13				0.18	1.00	1.13	0.88
	PK	21.62	0.45	22.02	0.18	0.00	0.25	1.00	0.18	0.12
BTC, n=5	JH	22.56	0.51							
	LB	19.34	0.41							
	MT	16.15	0.19							
	PS	18.00	0.16				0.14	10.91	12.03	9.89
	RB	16.84	0.04	18.58	0.14	-3.45	0.23	10.91	1.73	1.07
		ΔC_T (COL1A1-18S)		ΔC_T (COL1A1-18S)		ΔC_T - $\Delta C_{T,cal}$		COL1A1 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	26.01								
	JP	22.84	0.38				0.36	1.00	1.28	0.78
	PK	21.74	0.61	23.62	0.36	0.00	0.51	1.00	0.35	0.25
BTC, n=5	JH	17.46	0.17							
	LB	16.68	0.04							
	MT	21.72	0.36							
	PS	20.52	0.39				0.11	21.59	23.36	19.96
	RB	19.56	0.10	19.19	0.11	-4.43	0.38	21.59	5.62	1.70
		ΔC_T (COL6A3-18S)		ΔC_T (COL6A3-18S)		ΔC_T - $\Delta C_{T,cal}$		COL6A3 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	29.10								
	JP	24.59	0.17				0.17	1.00	1.13	0.89
	PK	22.33		24.98	0.17	0.00	0.24	1.00	0.17	0.12
BTC, n=5	JH	18.58	0.17							
	LB	17.84	0.10							
	MT	21.67	0.21							
	PS	21.58	0.19				0.07	26.71	28.10	25.39
	RB	21.53	0.12	20.24	0.07	-4.74	0.19	26.71	3.46	1.35
		ΔC_T (CSPG4-18S)		ΔC_T (CSPG4-18S)		ΔC_T - $\Delta C_{T,cal}$		CSPG4 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal,	AS1	23.90								

n=4										
	DA	29.10								
	JP	28.95					#DIV/0!	1.00	#DIV/0!	#DIV/0!
	PK	22.36		26.08	#DIV/0!	0.00	#DIV/0!	1.00	#DIV/0!	#DIV/0!
BTC, n=5	JH	22.89	1.46							
	LB	17.50	0.09							
	MT	21.50	0.39							
	PS	24.62	1.11				0.38	19.52	25.46	14.97
	RB	22.45	0.38	21.79	0.38	-4.29	#DIV/0!	19.52	#DIV/0!	5.18
		ΔC_T (HOXA10-18S)		**ΔC_T (HOXA10-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**HOXA10 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	21.90	0.91							
	JP	23.43	0.38				0.49	1.00	1.41	0.71
	PK	22.36		22.90	0.49	0.00	0.70	1.00	0.48	0.34
BTC, n=5	JH	13.97	0.14							
	LB	17.42	0.13							
	MT	20.45	0.27							
	PS	18.37	0.17				0.09	66.96	71.13	63.03
	RB	13.95	0.24	16.83	0.09	-6.07	0.50	66.96	23.28	4.05
		ΔC_T (HOXB6-18S)		**ΔC_T (HOXB6-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**HOXB6 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	18.95	0.07							
	DA	21.96	0.14							
	JP	20.80	0.16				0.07	1.00	1.05	0.95
	PK	19.89	0.19	20.40	0.07	0.00	0.10	1.00	0.07	0.05
BTC, n=5	JH	19.15	0.11							
	LB	15.88	0.10							
	MT	18.85	0.19							
	PS	19.48	0.15				0.06	4.99	5.20	4.80
	RB	17.05	0.05	18.08	0.06	-2.32	0.09	4.99	0.32	0.20
		ΔC_T (ITGB8-18S)		**ΔC_T (ITGB8-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**ITGB8 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	20.74	0.14							
	DA	19.11	0.06							
	JP	20.31	0.23				0.15	1.00	1.11	0.90
	PK	18.06	0.53	19.55	0.15	0.00	0.21	1.00	0.15	0.10
BTC, n=5	JH	17.13	0.08							
	LB	16.98	0.29							
	MT	17.98	0.26							
	PS	19.35	0.38				0.11	3.99	4.31	3.69
	RB	16.34	0.10	17.56	0.11	-1.99	0.19	3.99	0.52	0.31
		ΔC_T (ITIH5-18S)		**ΔC_T (ITIH5-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**ITIH5 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	16.86								
	DA	25.17	0.14							
	JP	28.95	0.08				0.08	1.00	1.06	0.95
	PK	21.78		23.19	0.08	0.00	0.11	1.00	0.08	0.06

BTC, n=5	JH	17.66	0.08							
	LB	18.18	1.45							
	MT	16.60	1.14							
	PS	24.14	0.18				0.38	12.42	16.15	9.55
	RB	21.21	0.38	19.56	0.38	-3.63	0.39	12.42	3.34	3.26
		ΔC_T (KRAS-18S)		**ΔC_T (KRAS-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**KRAS rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	18.07	0.06							
	DA	17.67	0.06							
	JP	18.79	0.09				0.04	1.00	1.03	0.97
	PK	17.33	0.08	17.96	0.04	0.00	0.05	1.00	0.04	0.03
BTC, n=5	JH	18.28	0.05							
	LB	17.27	0.07							
	MT	18.42	0.13							
	PS	18.72	0.15				0.05	1.05	1.08	1.02
	RB	16.80	0.06	17.90	0.05	-0.07	0.06	1.05	0.04	0.03
		ΔC_T (LAMC2-18S)		**ΔC_T (LAMC2-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**LAMC2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	20.07	0.23							
	DA	17.79	0.24							
	JP	18.04	0.27				0.11	1.00	1.08	0.92
	PK	17.94	0.15	18.46	0.11	0.00	0.16	1.00	0.11	0.08
BTC, n=5	JH	17.50	0.15							
	LB	16.55	0.17							
	MT	16.39	0.19							
	PS	18.29	0.22				0.07	2.79	2.93	2.65
	RB	16.16	0.05	16.98	0.07	-1.48	0.13	2.79	0.26	0.14
		ΔC_T (LEF1-18S)		**ΔC_T (LEF1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**LEF1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	26.05								
	JP	23.16	0.22				0.22	1.00	1.16	0.86
	PK	22.37		23.87	0.22	0.00	0.31	1.00	0.21	0.15
BTC, n=5	JH	20.35	0.23							
	LB	19.66	0.43							
	MT	21.42	0.35							
	PS	22.95	0.25				0.13	7.13	7.80	6.52
	RB	20.79	0.04	21.04	0.13	-2.83	0.25	7.13	1.25	0.64
		ΔC_T (LIF-18S)		**ΔC_T (LIF-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**LIF rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.83								
	DA	20.75	0.31							
	JP	20.35	0.09				0.13	1.00	1.10	0.91
	PK	18.31	0.23	19.81	0.13	0.00	0.19	1.00	0.13	0.09
BTC, n=5	JH	15.90	0.08							
	LB	14.63	0.04							
	MT	17.56	0.13							
	PS	18.93	0.24				0.06	7.51	7.83	7.20

	RB	17.50	0.10	16.90	0.06	-2.91	0.15	7.51	0.76	0.31
		ΔC_T (LUM-18S)		**ΔC_T (LUM-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**LUM rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	23.60	0.24							
	JP	23.20					0.24	1.00	1.18	0.85
	PK	22.25		23.24	0.24	0.00	0.34	1.00	0.24	0.17
BTC, n=5	JH	24.08	0.16							
	LB	18.47	0.22							
	MT	21.55	0.45							
	PS	21.02	0.69				0.19	3.70	4.22	3.24
	RB	21.62	0.39	21.35	0.19	-1.89	0.31	3.70	0.79	0.49
		ΔC_T (MAML2-18S)		**ΔC_T (MAML2-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**MAML2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	20.75								
	DA	21.59	0.15							
	JP	22.05	0.81				0.35	1.00	1.28	0.78
	PK	21.66	0.67	21.51	0.35	0.00	0.50	1.00	0.35	0.25
BTC, n=5	JH	19.24	0.05							
	LB	18.23	0.04							
	MT	20.69	0.19							
	PS	20.83	0.33				0.08	3.87	4.10	3.65
	RB	18.80	0.16	19.56	0.08	-1.95	0.36	3.87	0.98	0.23
		ΔC_T (MAML3-18S)		**ΔC_T (MAML3-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**MAML3 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.71	0.18							
	DA	20.07	0.06							
	JP	21.19	0.29				0.09	1.00	1.06	0.94
	PK	19.13	0.11	20.02	0.09	0.00	0.13	1.00	0.09	0.06
BTC, n=5	JH	21.02	0.20							
	LB	19.66	0.32							
	MT	21.54	0.30							
	PS	21.37	0.22				0.11	0.70	0.76	0.65
	RB	19.07	0.05	20.53	0.11	0.51	0.14	0.70	0.07	0.05
		ΔC_T (MAPK1-18S)		**ΔC_T (MAPK1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**MAPK1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.72	0.14							
	DA	18.53	0.07							
	JP	19.68	0.16				0.12	1.00	1.09	0.92
	PK	18.20	0.45	19.04	0.12	0.00	0.18	1.00	0.12	0.09
BTC, n=5	JH	18.57	0.19							
	LB	17.62	0.07							
	MT	19.39	0.27							
	PS	19.71	0.15				0.07	1.60	1.68	1.52
	RB	16.52	0.04	18.36	0.07	-0.67	0.14	1.60	0.16	0.08
		ΔC_T (MCM4-18S)		**ΔC_T (MCM4-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**MCM4 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal,	AS1	20.80	0.06							

n=4										
	DA	20.62	0.15							
	JP	21.68	0.18				0.07	1.00	1.05	0.95
	PK	18.98	0.13	20.52	0.07	0.00	0.10	1.00	0.07	0.05
BTC, n=5	JH	21.36	0.08							
	LB	18.87	0.22							
	MT	20.09	0.46							
	PS	21.87	0.23				0.11	1.50	1.62	1.38
	RB	17.50	0.03	19.94	0.11	-0.58	0.13	1.50	0.14	0.12
		ΔC_T (MMP2-18S)		**ΔC_T (MMP2-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**MMP2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	24.41	3.46							
	JP	26.70	0.34				1.18	1.00	2.26	0.44
	PK	21.71	0.66	24.18	1.18	0.00	1.67	1.00	1.15	0.82
BTC, n=5	JH	18.95	0.06							
	LB	19.22	0.15							
	MT	23.07	0.33							
	PS	20.09	0.26				0.10	11.00	11.79	10.26
	RB	22.25	0.22	20.72	0.10	-3.46	1.18	11.00	9.01	0.76
		ΔC_T (MXRA5-18S)		**ΔC_T (MXRA5-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**MXRA5 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	26.06								
	JP	22.97	0.44				0.44	1.00	1.35	0.74
	PK	22.37		23.83	0.44	0.00	0.62	1.00	0.43	0.30
BTC, n=5	JH	24.01	0.05							
	LB	18.38	0.20							
	MT	18.92	0.39							
	PS	20.03	0.15				0.10	10.41	11.14	9.74
	RB	20.90	0.12	20.45	0.10	-3.38	0.45	10.41	3.23	0.70
		ΔC_T (MYC-18S)		**ΔC_T (MYC-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**MYC rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	20.84								
	DA	20.97	0.08							
	JP	20.56	0.09				0.05	1.00	1.03	0.97
	PK	18.63	0.07	20.25	0.05	0.00	0.07	1.00	0.05	0.03
BTC, n=5	JH	17.87	0.10							
	LB	16.03	0.05							
	MT	19.42	0.21							
	PS	19.55	0.18				0.06	5.96	6.22	5.72
	RB	15.48	0.02	17.67	0.06	-2.58	0.08	5.96	0.31	0.25
		ΔC_T (NCSTN-18S)		**ΔC_T (NCSTN-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**NCSTN rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	17.72	0.15							
	DA	18.02	0.06							
	JP	19.37	0.11				0.05	1.00	1.04	0.97
	PK	17.34	0.07	18.11	0.05	0.00	0.07	1.00	0.05	0.04

BTC, n=5	JH	18.10	0.18							
	LB	17.90	0.14							
	MT	18.89	0.24							
	PS	18.80	0.16				0.07	0.95	1.00	0.90
	RB	17.23	0.02	18.18	0.07	0.07	0.09	0.95	0.06	0.05
		ΔC_T (PEAR1-18S)		**ΔC_T (PEAR1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**PEAR1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	23.14	0.54							
	JP	25.90					0.54	1.00	1.45	0.69
	PK	21.35		23.57	0.54	0.00	0.76	1.00	0.53	0.37
BTC, n=5	JH	18.91	0.10							
	LB	21.00	0.04							
	MT	22.24	0.21							
	PS	21.54	0.29				0.09	6.61	7.02	6.24
	RB	20.55	0.20	20.85	0.09	-2.73	0.54	6.61	2.49	0.39
		ΔC_T (PNMA2-18S)		**ΔC_T (PNMA2-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**PNMA2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.97	0.84							
	DA	23.54	0.06							
	JP	20.49	0.21				0.22	1.00	1.16	0.86
	PK	25.40	0.07	22.35	0.22	0.00	0.31	1.00	0.21	0.15
BTC, n=5	JH	24.08	0.21							
	LB	17.22	0.36							
	MT	20.50	0.13							
	PS	22.24	0.52				0.14	3.89	4.28	3.53
	RB	17.92	0.10	20.39	0.14	-1.96	0.26	3.89	0.69	0.37
		ΔC_T (POU5F1-18S)		**ΔC_T (POU5F1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**POU5F1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.41	0.46							
	DA	19.98	0.10							
	JP	20.39	0.08				0.12	1.00	1.09	0.92
	PK	20.01	0.09	19.95	0.12	0.00	0.17	1.00	0.12	0.08
BTC, n=5	JH	19.93	0.10							
	LB	17.97	0.30							
	MT	19.18	0.15							
	PS	19.54	0.20				0.08	2.52	2.66	2.38
	RB	16.45	0.05	18.61	0.08	-1.33	0.15	2.52	0.25	0.14
		ΔC_T (PRKCB1-18S)		**ΔC_T (PRKCB1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**PRKCB1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.97	0.11							
	DA	24.99	1.07							
	JP	21.80	0.48				0.30	1.00	1.23	0.81
	PK	19.27	0.26	21.51	0.30	0.00	0.43	1.00	0.30	0.21
BTC, n=5	JH	18.46	0.06							
	LB	16.72	0.13							
	MT	19.80	0.16							

	PS	18.44	0.20				0.06	6.81	7.13	6.51
	RB	20.28	0.15	18.74	0.06	-2.77	0.31	6.81	1.46	0.31
		ΔC_T (PTK2-18S)		**ΔC_T (PTK2-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**PTK2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	18.31	0.45							
	DA	18.03	0.07							
	JP	19.73	0.14				0.12	1.00	1.09	0.92
	PK	16.89	0.08	18.24	0.12	0.00	0.17	1.00	0.12	0.08
BTC, n=5	JH	16.57	0.10							
	LB	16.93	0.11							
	MT	18.59	0.13							
	PS	18.37	0.15				0.05	1.94	2.01	1.87
	RB	15.95	0.08	17.28	0.05	-0.95	0.13	1.94	0.18	0.07
		ΔC_T (PVT1 -18S)		**ΔC_T (PVT1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**PVT1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	24.00	2.09							
	JP	21.54	0.18				1.05	1.00	2.07	0.48
	PK	22.35		22.95	1.05	0.00	1.49	1.00	1.03	0.73
BTC, n=5	JH	17.10	0.08							
	LB	21.54	0.17							
	MT	22.06	0.57							
	PS	18.69	0.20				0.13	12.07	13.17	11.05
	RB	17.39	0.03	19.35	0.13	-3.59	1.06	12.07	8.85	1.06
		ΔC_T (RAB27A-18S)		**ΔC_T (RAB27A-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**RAB27A rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	20.25	0.45							
	DA	20.11	0.09							
	JP	20.13	0.20				0.14	1.00	1.10	0.91
	PK	19.54	0.22	20.01	0.14	0.00	0.19	1.00	0.13	0.10
BTC, n=5	JH	19.78	0.05							
	LB	17.72	0.10							
	MT	18.57	0.17							
	PS	20.00	0.19				0.06	2.74	2.85	2.62
	RB	16.71	0.12	18.56	0.06	-1.45	0.15	2.74	0.28	0.12
		ΔC_T (RAD51-18S)		**ΔC_T (RAD51-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**RAD51 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	23.73	0.41							
	JP	22.87	0.10				0.21	1.00	1.16	0.87
	PK	21.37		22.97	0.21	0.00	0.29	1.00	0.20	0.14
BTC, n=5	JH	22.87								
	LB	20.76	0.46							
	MT	20.82	0.35							
	PS	24.37	0.16				0.16	2.98	3.33	2.67
	RB	18.14	0.23	21.39	0.16	-1.58	0.26	2.98	0.54	0.33
		ΔC_T (SERPINA3-18S)		**ΔC_T (SERPINA3-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**SERPINA3 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	

Normal, n=4	AS1	23.90								
	DA	23.65	0.13							
	JP	17.77	0.13				0.15	1.00	1.11	0.90
	PK	21.56	0.40	21.72	0.15	0.00	0.21	1.00	0.14	0.10
BTC, n=5	JH	13.77	0.09							
	LB	14.75	0.08							
	MT	17.10	0.13							
	PS	16.54	0.15				0.05	46.74	48.30	45.23
	RB	18.71	0.03	16.17	0.05	-5.55	0.16	46.74	5.03	1.53
		ΔC_T (SPARC-18S)		**ΔC_T (SPARC-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**SPARC rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	29.10								
	JP	24.92	0.98				0.98	1.00	1.97	0.51
	PK	22.37		25.07	0.98	0.00	1.39	1.00	0.96	0.68
BTC, n=5	JH	18.80	0.19							
	LB	17.51	0.22							
	MT	20.90	0.48							
	PS	20.69	0.17				0.12	41.37	45.04	37.99
	RB	20.62	0.17	19.70	0.12	-5.37	0.99	41.37	28.35	3.52
		ΔC_T (STAT1-18S)		**ΔC_T (STAT1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**STAT1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	17.45	0.17							
	DA	17.02	0.06							
	JP	18.76	0.19				0.07	1.00	1.05	0.95
	PK	16.11	0.07	17.33	0.07	0.00	0.10	1.00	0.07	0.05
BTC, n=5	JH	16.81	0.20							
	LB	16.49	0.06							
	MT	16.78	0.14							
	PS	17.16	0.15				0.06	1.91	1.99	1.83
	RB	14.76	0.12	16.40	0.06	-0.93	0.09	1.91	0.12	0.08
		ΔC_T (TGFBI-18S)		**ΔC_T (TGFBI-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**TGFBI rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.75	0.16							
	DA	21.79	0.07							
	JP	21.66	0.14				0.11	1.00	1.08	0.93
	PK	18.82	0.39	20.50	0.11	0.00	0.16	1.00	0.11	0.08
BTC, n=5	JH	18.87	0.06							
	LB	17.13	0.18							
	MT	18.44	0.24							
	PS	20.72	0.15				0.07	3.28	3.45	3.12
	RB	18.79	0.11	18.79	0.07	-1.71	0.13	3.28	0.30	0.16
		ΔC_T (TM4SF18-18S)		**ΔC_T (TM4SF18-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**TM4SF18 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	8.63								
	DA	24.16	0.35							
	JP	25.89					0.48	1.00	1.40	0.72

	PK	15.39	0.90	18.52	0.48	0.00	0.68	1.00	0.47	0.33
BTC, n=5	JH	23.74	0.47							
	LB	16.42	0.63							
	MT	18.06								
	PS	20.39	0.15				0.20	0.93	1.06	0.81
	RB	14.53	0.03	18.63	0.20	0.11	0.52	0.93	0.34	0.13
		ΔC_T (TRIB2-18S)		**ΔC_T (TRIB2-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**TRIB2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.15	0.72							
	DA	19.05	0.06							
	JP	19.56	0.26				0.19	1.00	1.14	0.87
	PK	17.46	0.13	18.81	0.19	0.00	0.27	1.00	0.19	0.13
BTC, n=5	JH	17.44	0.29							
	LB	15.89	0.30							
	MT	17.85	0.14							
	PS	18.02	0.15				0.09	3.39	3.61	3.18
	RB	16.03	0.03	17.05	0.09	-1.76	0.21	3.39	0.50	0.22
		ΔC_T (VEGFA-18S)		**ΔC_T (VEGFA-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**VEGFA rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	17.36	0.19							
	DA	17.00	0.11							
	JP	18.09	0.08				0.08	1.00	1.06	0.95
	PK	16.67		17.28	0.08	0.00	0.11	1.00	0.08	0.05
BTC, n=5	JH	15.83	0.05							
	LB	15.44	0.13							
	MT	17.13	0.14							
	PS	16.11					0.05	2.62	2.71	2.53
	RB	14.95	0.04	15.89	0.05	-1.39	0.09	2.62	0.17	0.09
						$\Delta\Delta C_T$				
		ΔC_T (GAPDH-18S)		**ΔC_T (GAPDH-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**GAPDH rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	15.07	0.12							
	DA	13.69	0.19							
	JP	14.45	0.08				0.07	1.00	1.05	0.95
	PK	13.23	0.18	14.11	0.07	0.00	0.10	1.00	0.07	0.05
BTC, n=5	JH	14.08	0.25							
	LB	13.60	0.04							
	MT	15.22	0.30							
	PS	14.70	0.15				0.09	1.00	1.06	0.94
	RB	12.92	0.06	14.11	0.09	0.00	0.11	1.00	0.08	0.06
						$\Delta\Delta C_T$				
		ΔC_T (18S-18S)		**ΔC_T (18S-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**18S rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1									
	DA									
	JP						#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
	PK			#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!

BTC, n=5	JH									
	LB									
	MT									
	PS						0.09	#DIV/0!	#DIV/0!	#DIV/0!
	RB			#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
						$\Delta\Delta C_T$				
		ΔC_T (MUC4-18S)		ΔC_T (MUC4-18S)		$\Delta C_T - \Delta C_{T,cal}$		MUC4 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	26.03								
	JP	20.88	0.45				0.35	1.00	1.27	0.79
	PK	21.85	0.53	23.16	0.35	0.00	0.49	1.00	0.34	0.24
BTC, n=5	JH	19.95	0.09							
	LB	16.11	0.05							
	MT	17.83	0.53							
	PS	18.44	0.15				0.09	29.37	31.18	27.66
	RB	19.10	0.15	18.29	0.12	-4.88	0.37	29.37	7.47	2.34
						$\Delta\Delta C_T$				
		ΔC_T (MUC5AC-18S)		ΔC_T (MUC5AC-18S)		$\Delta C_T - \Delta C_{T,cal}$		MUC5AC rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	18.50	1.40							
	DA	18.05	0.08							
	JP	13.84	0.11				0.36	1.00	1.28	0.78
	PK	13.60	0.24	16.00	0.36	0.00	0.50	1.00	0.35	0.25
BTC, n=5	JH	14.40	0.09							
	LB	12.33	0.12							
	MT	10.38	0.18							
	PS	15.94	0.22				0.06	7.41	7.75	7.09
	RB	12.49	0.07	13.11	0.06	-2.89	0.36	7.41	1.86	0.33
						$\Delta\Delta C_T$				
		ΔC_T (ACTB-18S)		ΔC_T (ACTB-18S)		$\Delta C_T - \Delta C_{T,cal}$		ACTB rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	12.02	0.12							
	DA	13.03	0.07							
	JP	12.93	0.08				0.04	1.00	1.03	0.97
	PK	11.40	0.08	12.34	0.04	0.00	0.06	1.00	0.04	0.03
BTC, n=5	JH	11.25	0.06							
	LB	10.29	0.20							
	MT	12.49	0.13							
	PS	12.55	0.15				0.06	2.03	2.11	1.95
	RB	10.03	0.02	11.32	0.06	-1.02	0.07	2.03	0.10	0.08
						$\Delta\Delta C_T$				
		ΔC_T (CD9-18S)		ΔC_T (CD9-18S)		$\Delta C_T - \Delta C_{T,cal}$		CD9 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	18.21	0.38							
	DA	18.00	0.07							
	JP	17.05	0.08				0.11	1.00	1.08	0.93
	PK	17.07	0.16	17.58	0.11	0.00	0.15	1.00	0.11	0.07
BTC,	JH	16.98	0.12							

n=5										
	LB	14.99	0.06							
	MT	16.86	0.15							
	PS	18.09	0.16				0.05	2.15	2.23	2.07
	RB	15.48	0.07	16.48	0.05	-1.10	0.12	2.15	0.18	0.08

		ΔC_T (Notch3-18S)		ΔC_T (Notch3-18S)		$\Delta\Delta C_T$ $\Delta C_T - \Delta C_{T,cal}$		Notch3 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	23.79	0.26							
	JP	21.45	0.37				0.18	1.00	1.13	0.88
	PK	21.05	0.30	22.55	0.18	0.00	0.25	1.00	0.18	0.12
BTC, n=5	JH	21.02	0.14							
	LB	17.36	0.13							
	MT	20.68	0.32							
	PS	21.41	0.19				0.16	8.01	8.96	7.16
	RB	17.27	0.69	19.55	0.16	-3.00	0.24	8.01	1.34	0.90

		ΔC_T (ASPHD1-18S)		ΔC_T (ASPHD1-18S)		$\Delta\Delta C_T$ $\Delta C_T - \Delta C_{T,cal}$		ASPDH1 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.88	0.98							
	DA	18.32	0.10							
	JP	20.47	0.18				0.25	1.00	1.19	0.84
	PK	17.58	0.07	19.06	0.25	0.00	0.35	1.00	0.25	0.17
BTC, n=5	JH	19.21	0.09							
	LB	17.02	0.33							
	MT	18.73	0.14							
	PS	20.56	0.18				0.08	1.65	1.75	1.56
	RB	16.18	0.02	18.34	0.08	-0.72	0.26	1.65	0.30	0.09

		ΔC_T (ATP6V0A2-18S)		ΔC_T (ATP6VOA2-18S)		$\Delta\Delta C_T$ $\Delta C_T - \Delta C_{T,cal}$		ATP6V0A2 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.56	0.29							
	DA	19.97	0.10							
	JP	21.12	0.36				0.12	1.00	1.09	0.92
	PK	19.24	0.15	19.97	0.12	0.00	0.17	1.00	0.12	0.08
BTC, n=5	JH	19.75	0.08							
	LB	18.71	0.17							
	MT	20.50	0.40							
	PS	21.38	0.23				0.11	1.10	1.18	1.02
	RB	18.84	0.17	19.84	0.11	-0.14	0.16	1.10	0.12	0.08

		ΔC_T (CEACAM1-18S)		ΔC_T (CEACAM1-18S)		$\Delta C_T - \Delta C_{T,cal}$		CEACAM1 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	18.25	0.05							
	DA	20.31	0.06							
	JP	20.86	0.08				0.09	1.00	1.06	0.94
	PK	20.02	0.32	19.86	0.09	0.00	0.12	1.00	0.08	0.06
BTC,	JH	20.60	0.11							

n=5										
	LB	18.28	0.04							
	MT	19.42	0.13							
	PS	19.05	0.17				0.05	2.08	2.16	2.00
	RB	16.67	0.12	18.80	0.05	-1.05	0.10	2.08	0.15	0.08
		ΔC_T (CELSR1-18S)		**ΔC_T (CELSR1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**CELSR1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	24.28	0.27							
	JP	22.94	0.08				0.11	1.00	1.08	0.93
	PK	21.56	0.18	23.17	0.11	0.00	0.16	1.00	0.11	0.08
BTC, n=5	JH	20.85	0.58							
	LB	19.37	0.11							
	MT	19.72	0.40							
	PS	23.09	0.17				0.15	7.15	7.92	6.45
	RB	18.64	0.08	20.33	0.15	-2.84	0.19	7.15	0.92	0.73
		ΔC_T (CFDP1 - 18S)		**ΔC_T (CFDP1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**CFDP1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.05	0.79							
	DA	19.02	0.07							
	JP	20.33	0.11				0.21	1.00	1.15	0.87
	PK	17.14	0.24	18.88	0.21	0.00	0.29	1.00	0.20	0.14
BTC, n=5	JH	19.16	0.21							
	LB	17.12	0.06							
	MT	19.40	0.13							
	PS	20.19	0.20				0.07	1.11	1.16	1.06
	RB	17.78	0.03	18.73	0.07	-0.15	0.22	1.11	0.17	0.05
		ΔC_T (COL17A1-18S)		**ΔC_T (COL17A1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**COL17A1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.88								
	DA	25.32	0.27							
	JP	21.28	0.13				0.18	1.00	1.13	0.88
	PK	21.62	0.45	22.02	0.18	0.00	0.25	1.00	0.18	0.12
BTC, n=5	JH	22.56	0.51							
	LB	19.34	0.41							
	MT	16.15	0.19							
	PS	18.00	0.16				0.14	10.91	12.03	9.89
	RB	16.84	0.04	18.58	0.14	-3.45	0.23	10.91	1.73	1.07
		ΔC_T (COL1A1-18S)		**ΔC_T (COL1A1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**COL1A1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	26.01								
	JP	22.84	0.38				0.36	1.00	1.28	0.78
	PK	21.74	0.61	23.62	0.36	0.00	0.51	1.00	0.35	0.25
BTC, n=5	JH	17.46	0.17							
	LB	16.68	0.04							
	MT	21.72	0.36							

	PS	20.52	0.39				0.11	21.59	23.36	19.96
	RB	19.56	0.10	19.19	0.11	-4.43	0.38	21.59	5.62	1.70
		ΔC_T (COL6A3-18S)		**ΔC_T (COL6A3-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**COL6A3 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	29.10								
	JP	24.59	0.17				0.17	1.00	1.13	0.89
	PK	22.33		24.98	0.17	0.00	0.24	1.00	0.17	0.12
BTC, n=5	JH	18.58	0.17							
	LB	17.84	0.10							
	MT	21.67	0.21							
	PS	21.58	0.19				0.07	26.71	28.10	25.39
	RB	21.53	0.12	20.24	0.07	-4.74	0.19	26.71	3.46	1.35
		ΔC_T (CSPG4-18S)		**ΔC_T (CSPG4-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**CSPG4 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	29.10								
	JP	28.95					#DIV/0!	1.00	#DIV/0!	#DIV/0!
	PK	22.36		26.08	#DIV/0!	0.00	#DIV/0!	1.00	#DIV/0!	#DIV/0!
BTC, n=5	JH	22.89	1.46							
	LB	17.50	0.09							
	MT	21.50	0.39							
	PS	24.62	1.11				0.38	19.52	25.46	14.97
	RB	22.45	0.38	21.79	0.38	-4.29	#DIV/0!	19.52	#DIV/0!	5.18
		ΔC_T (HOXA10-18S)		**ΔC_T (HOXA10-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**HOXA10 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	21.90	0.91							
	JP	23.43	0.38				0.49	1.00	1.41	0.71
	PK	22.36		22.90	0.49	0.00	0.70	1.00	0.48	0.34
BTC, n=5	JH	13.97	0.14							
	LB	17.42	0.13							
	MT	20.45	0.27							
	PS	18.37	0.17				0.09	66.96	71.13	63.03
	RB	13.95	0.24	16.83	0.09	-6.07	0.50	66.96	23.28	4.05
		ΔC_T (HOXB6-18S)		**ΔC_T (HOXB6-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**HOXB6 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	18.95	0.07							
	DA	21.96	0.14							
	JP	20.80	0.16				0.07	1.00	1.05	0.95
	PK	19.89	0.19	20.40	0.07	0.00	0.10	1.00	0.07	0.05
BTC, n=5	JH	19.15	0.11							
	LB	15.88	0.10							
	MT	18.85	0.19							
	PS	19.48	0.15				0.06	4.99	5.20	4.80
	RB	17.05	0.05	18.08	0.06	-2.32	0.09	4.99	0.32	0.20
		ΔC_T (ITGB8-18S)		**ΔC_T (ITGB8-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**ITGB8 rel to Cal**		

	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	20.74	0.14							
	DA	19.11	0.06							
	JP	20.31	0.23				0.15	1.00	1.11	0.90
	PK	18.06	0.53	19.55	0.15	0.00	0.21	1.00	0.15	0.10
BTC, n=5	JH	17.13	0.08							
	LB	16.98	0.29							
	MT	17.98	0.26							
	PS	19.35	0.38				0.11	3.99	4.31	3.69
	RB	16.34	0.10	17.56	0.11	-1.99	0.19	3.99	0.52	0.31
		ΔC_T (ITIH5-18S)		**ΔC_T (ITIH5-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**ITIH5 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	16.86								
	DA	25.17	0.14							
	JP	28.95	0.08				0.08	1.00	1.06	0.95
	PK	21.78		23.19	0.08	0.00	0.11	1.00	0.08	0.06
BTC, n=5	JH	17.66	0.08							
	LB	18.18	1.45							
	MT	16.60	1.14							
	PS	24.14	0.18				0.38	12.42	16.15	9.55
	RB	21.21	0.38	19.56	0.38	-3.63	0.39	12.42	3.34	3.26
		ΔC_T (KRAS-18S)		**ΔC_T (KRAS-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**KRAS rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	18.07	0.06							
	DA	17.67	0.06							
	JP	18.79	0.09				0.04	1.00	1.03	0.97
	PK	17.33	0.08	17.96	0.04	0.00	0.05	1.00	0.04	0.03
BTC, n=5	JH	18.28	0.05							
	LB	17.27	0.07							
	MT	18.42	0.13							
	PS	18.72	0.15				0.05	1.05	1.08	1.02
	RB	16.80	0.06	17.90	0.05	-0.07	0.06	1.05	0.04	0.03
		ΔC_T (LAMC2-18S)		**ΔC_T (LAMC2-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**LAMC2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	20.07	0.23							
	DA	17.79	0.24							
	JP	18.04	0.27				0.11	1.00	1.08	0.92
	PK	17.94	0.15	18.46	0.11	0.00	0.16	1.00	0.11	0.08
BTC, n=5	JH	17.50	0.15							
	LB	16.55	0.17							
	MT	16.39	0.19							
	PS	18.29	0.22				0.07	2.79	2.93	2.65
	RB	16.16	0.05	16.98	0.07	-1.48	0.13	2.79	0.26	0.14
		ΔC_T (LEF1-18S)		**ΔC_T (LEF1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**LEF1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	26.05								
	JP	23.16	0.22				0.22	1.00	1.16	0.86

	PK	22.37		23.87	0.22	0.00	0.31	1.00	0.21	0.15
BTC, n=5	JH	20.35	0.23							
	LB	19.66	0.43							
	MT	21.42	0.35							
	PS	22.95	0.25				0.13	7.13	7.80	6.52
	RB	20.79	0.04	21.04	0.13	-2.83	0.25	7.13	1.25	0.64
		ΔC_T (LIF-18S)		**ΔC_T (LIF-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**LIF rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.83								
	DA	20.75	0.31							
	JP	20.35	0.09				0.13	1.00	1.10	0.91
	PK	18.31	0.23	19.81	0.13	0.00	0.19	1.00	0.13	0.09
BTC, n=5	JH	15.90	0.08							
	LB	14.63	0.04							
	MT	17.56	0.13							
	PS	18.93	0.24				0.06	7.51	7.83	7.20
	RB	17.50	0.10	16.90	0.06	-2.91	0.15	7.51	0.76	0.31
		ΔC_T (LUM-18S)		**ΔC_T (LUM-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**LUM rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	23.60	0.24							
	JP	23.20					0.24	1.00	1.18	0.85
	PK	22.25		23.24	0.24	0.00	0.34	1.00	0.24	0.17
BTC, n=5	JH	24.08	0.16							
	LB	18.47	0.22							
	MT	21.55	0.45							
	PS	21.02	0.69				0.19	3.70	4.22	3.24
	RB	21.62	0.39	21.35	0.19	-1.89	0.31	3.70	0.79	0.49
		ΔC_T (MAML2-18S)		**ΔC_T (MAML2-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**MAML2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	20.75								
	DA	21.59	0.15							
	JP	22.05	0.81				0.35	1.00	1.28	0.78
	PK	21.66	0.67	21.51	0.35	0.00	0.50	1.00	0.35	0.25
BTC, n=5	JH	19.24	0.05							
	LB	18.23	0.04							
	MT	20.69	0.19							
	PS	20.83	0.33				0.08	3.87	4.10	3.65
	RB	18.80	0.16	19.56	0.08	-1.95	0.36	3.87	0.98	0.23
		ΔC_T (MAML3-18S)		**ΔC_T (MAML3-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**MAML3 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.71	0.18							
	DA	20.07	0.06							
	JP	21.19	0.29				0.09	1.00	1.06	0.94
	PK	19.13	0.11	20.02	0.09	0.00	0.13	1.00	0.09	0.06
BTC, n=5	JH	21.02	0.20							
	LB	19.66	0.32							
	MT	21.54	0.30							

	PS	21.37	0.22				0.11	0.70	0.76	0.65
	RB	19.07	0.05	20.53	0.11	0.51	0.14	0.70	0.07	0.05
		ΔC_T (MAPK1-18S)		**ΔC_T (MAPK1-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**MAPK1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.72	0.14							
	DA	18.53	0.07							
	JP	19.68	0.16				0.12	1.00	1.09	0.92
	PK	18.20	0.45	19.04	0.12	0.00	0.18	1.00	0.12	0.09
BTC, n=5	JH	18.57	0.19							
	LB	17.62	0.07							
	MT	19.39	0.27							
	PS	19.71	0.15				0.07	1.60	1.68	1.52
	RB	16.52	0.04	18.36	0.07	-0.67	0.14	1.60	0.16	0.08
		ΔC_T (MCM4-18S)		**ΔC_T (MCM4-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**MCM4 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	20.80	0.06							
	DA	20.62	0.15							
	JP	21.68	0.18				0.07	1.00	1.05	0.95
	PK	18.98	0.13	20.52	0.07	0.00	0.10	1.00	0.07	0.05
BTC, n=5	JH	21.36	0.08							
	LB	18.87	0.22							
	MT	20.09	0.46							
	PS	21.87	0.23				0.11	1.50	1.62	1.38
	RB	17.50	0.03	19.94	0.11	-0.58	0.13	1.50	0.14	0.12
		ΔC_T (MMP2-18S)		**ΔC_T (MMP2-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**MMP2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	24.41	3.46							
	JP	26.70	0.34				1.18	1.00	2.26	0.44
	PK	21.71	0.66	24.18	1.18	0.00	1.67	1.00	1.15	0.82
BTC, n=5	JH	18.95	0.06							
	LB	19.22	0.15							
	MT	23.07	0.33							
	PS	20.09	0.26				0.10	11.00	11.79	10.26
	RB	22.25	0.22	20.72	0.10	-3.46	1.18	11.00	9.01	0.76
		ΔC_T (MXRA5-18S)		**ΔC_T (MXRA5-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**MXRA5 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	26.06								
	JP	22.97	0.44				0.44	1.00	1.35	0.74
	PK	22.37		23.83	0.44	0.00	0.62	1.00	0.43	0.30
BTC, n=5	JH	24.01	0.05							
	LB	18.38	0.20							
	MT	18.92	0.39							
	PS	20.03	0.15				0.10	10.41	11.14	9.74
	RB	20.90	0.12	20.45	0.10	-3.38	0.45	10.41	3.23	0.70
		ΔC_T (MYC-18S)		**ΔC_T (MYC-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**MYC rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	

Normal, n=4	AS1	20.84								
	DA	20.97	0.08							
	JP	20.56	0.09				0.05	1.00	1.03	0.97
	PK	18.63	0.07	20.25	0.05	0.00	0.07	1.00	0.05	0.03
BTC, n=5	JH	17.87	0.10							
	LB	16.03	0.05							
	MT	19.42	0.21							
	PS	19.55	0.18				0.06	5.96	6.22	5.72
	RB	15.48	0.02	17.67	0.06	-2.58	0.08	5.96	0.31	0.25
		ΔC_T (NCSTN-18S)		**ΔC_T (NCSTN-18S)**		**ΔC_T -$\Delta C_{T,cal}$**		**NCSTN rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	17.72	0.15							
	DA	18.02	0.06							
	JP	19.37	0.11				0.05	1.00	1.04	0.97
	PK	17.34	0.07	18.11	0.05	0.00	0.07	1.00	0.05	0.04
BTC, n=5	JH	18.10	0.18							
	LB	17.90	0.14							
	MT	18.89	0.24							
	PS	18.80	0.16				0.07	0.95	1.00	0.90
	RB	17.23	0.02	18.18	0.07	0.07	0.09	0.95	0.06	0.05
		ΔC_T (PEAR1-18S)		**ΔC_T (PEAR1-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**PEAR1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	23.14	0.54							
	JP	25.90					0.54	1.00	1.45	0.69
	PK	21.35		23.57	0.54	0.00	0.76	1.00	0.53	0.37
BTC, n=5	JH	18.91	0.10							
	LB	21.00	0.04							
	MT	22.24	0.21							
	PS	21.54	0.29				0.09	6.61	7.02	6.24
	RB	20.55	0.20	20.85	0.09	-2.73	0.54	6.61	2.49	0.39
		ΔC_T (PNMA2-18S)		**ΔC_T (PNMA2-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**PNMA2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.97	0.84							
	DA	23.54	0.06							
	JP	20.49	0.21				0.22	1.00	1.16	0.86
	PK	25.40	0.07	22.35	0.22	0.00	0.31	1.00	0.21	0.15
BTC, n=5	JH	24.08	0.21							
	LB	17.22	0.36							
	MT	20.50	0.13							
	PS	22.24	0.52				0.14	3.89	4.28	3.53
	RB	17.92	0.10	20.39	0.14	-1.96	0.26	3.89	0.69	0.37
		ΔC_T (POU5F1-18S)		**ΔC_T (POU5F1-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**POU5F1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.41	0.46							
	DA	19.98	0.10							
	JP	20.39	0.08				0.12	1.00	1.09	0.92

	PK	20.01	0.09	19.95	0.12	0.00	0.17	1.00	0.12	0.08
BTC, n=5	JH	19.93	0.10							
	LB	17.97	0.30							
	MT	19.18	0.15							
	PS	19.54	0.20				0.08	2.52	2.66	2.38
	RB	16.45	0.05	18.61	0.08	-1.33	0.15	2.52	0.25	0.14
		ΔC_T (PRKCB1-18S)		**ΔC_T (PRKCB1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**PRKCB1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.97	0.11							
	DA	24.99	1.07							
	JP	21.80	0.48				0.30	1.00	1.23	0.81
	PK	19.27	0.26	21.51	0.30	0.00	0.43	1.00	0.30	0.21
BTC, n=5	JH	18.46	0.06							
	LB	16.72	0.13							
	MT	19.80	0.16							
	PS	18.44	0.20				0.06	6.81	7.13	6.51
	RB	20.28	0.15	18.74	0.06	-2.77	0.31	6.81	1.46	0.31
		ΔC_T (PTK2-18S)		**ΔC_T (PTK2-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**PTK2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	18.31	0.45							
	DA	18.03	0.07							
	JP	19.73	0.14				0.12	1.00	1.09	0.92
	PK	16.89	0.08	18.24	0.12	0.00	0.17	1.00	0.12	0.08
BTC, n=5	JH	16.57	0.10							
	LB	16.93	0.11							
	MT	18.59	0.13							
	PS	18.37	0.15				0.05	1.94	2.01	1.87
	RB	15.95	0.08	17.28	0.05	-0.95	0.13	1.94	0.18	0.07
		ΔC_T (PVT1 -18S)		**ΔC_T (PVT1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**PVT1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	24.00	2.09							
	JP	21.54	0.18				1.05	1.00	2.07	0.48
	PK	22.35		22.95	1.05	0.00	1.49	1.00	1.03	0.73
BTC, n=5	JH	17.10	0.08							
	LB	21.54	0.17							
	MT	22.06	0.57							
	PS	18.69	0.20				0.13	12.07	13.17	11.05
	RB	17.39	0.03	19.35	0.13	-3.59	1.06	12.07	8.85	1.06
		ΔC_T (RAB27A-18S)		**ΔC_T (RAB27A-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**RAB27A rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	20.25	0.45							
	DA	20.11	0.09							
	JP	20.13	0.20				0.14	1.00	1.10	0.91
	PK	19.54	0.22	20.01	0.14	0.00	0.19	1.00	0.13	0.10
BTC, n=5	JH	19.78	0.05							
	LB	17.72	0.10							
	MT	18.57	0.17							

	PS	20.00	0.19				0.06	2.74	2.85	2.62
	RB	16.71	0.12	18.56	0.06	-1.45	0.15	2.74	0.28	0.12
		ΔC_T (RAD51-18S)		**ΔC_T (RAD51-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**RAD51 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	23.73	0.41							
	JP	22.87	0.10				0.21	1.00	1.16	0.87
	PK	21.37		22.97	0.21	0.00	0.29	1.00	0.20	0.14
BTC, n=5	JH	22.87								
	LB	20.76	0.46							
	MT	20.82	0.35							
	PS	24.37	0.16				0.16	2.98	3.33	2.67
	RB	18.14	0.23	21.39	0.16	-1.58	0.26	2.98	0.54	0.33
		ΔC_T (SERPINA3-18S)		**ΔC_T (SERPINA3-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**SERPINA3 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	23.65	0.13							
	JP	17.77	0.13				0.15	1.00	1.11	0.90
	PK	21.56	0.40	21.72	0.15	0.00	0.21	1.00	0.14	0.10
BTC, n=5	JH	13.77	0.09							
	LB	14.75	0.08							
	MT	17.10	0.13							
	PS	16.54	0.15				0.05	46.74	48.30	45.23
	RB	18.71	0.03	16.17	0.05	-5.55	0.16	46.74	5.03	1.53
		ΔC_T (SPARC-18S)		**ΔC_T (SPARC-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**SPARC rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	23.90								
	DA	29.10								
	JP	24.92	0.98				0.98	1.00	1.97	0.51
	PK	22.37		25.07	0.98	0.00	1.39	1.00	0.96	0.68
BTC, n=5	JH	18.80	0.19							
	LB	17.51	0.22							
	MT	20.90	0.48							
	PS	20.69	0.17				0.12	41.37	45.04	37.99
	RB	20.62	0.17	19.70	0.12	-5.37	0.99	41.37	28.35	3.52
		ΔC_T (STAT1-18S)		**ΔC_T (STAT1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**STAT1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	17.45	0.17							
	DA	17.02	0.06							
	JP	18.76	0.19				0.07	1.00	1.05	0.95
	PK	16.11	0.07	17.33	0.07	0.00	0.10	1.00	0.07	0.05
BTC, n=5	JH	16.81	0.20							
	LB	16.49	0.06							
	MT	16.78	0.14							
	PS	17.16	0.15				0.06	1.91	1.99	1.83
	RB	14.76	0.12	16.40	0.06	-0.93	0.09	1.91	0.12	0.08
		ΔC_T (TGFBI-18S)		**ΔC_T (TGFBI-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**TGFBI rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	

Normal, n=4	AS1	19.75	0.16							
	DA	21.79	0.07							
	JP	21.66	0.14				0.11	1.00	1.08	0.93
	PK	18.82	0.39	20.50	0.11	0.00	0.16	1.00	0.11	0.08
BTC, n=5	JH	18.87	0.06							
	LB	17.13	0.18							
	MT	18.44	0.24							
	PS	20.72	0.15				0.07	3.28	3.45	3.12
	RB	18.79	0.11	18.79	0.07	-1.71	0.13	3.28	0.30	0.16
		ΔC_T (TM4SF18-18S)		**ΔC_T (TM4SF18-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**TM4SF18 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	8.63								
	DA	24.16	0.35							
	JP	25.89					0.48	1.00	1.40	0.72
	PK	15.39	0.90	18.52	0.48	0.00	0.68	1.00	0.47	0.33
BTC, n=5	JH	23.74	0.47							
	LB	16.42	0.63							
	MT	18.06								
	PS	20.39	0.15				0.20	0.93	1.06	0.81
	RB	14.53	0.03	18.63	0.20	0.11	0.52	0.93	0.34	0.13
		ΔC_T (TRIB2-18S)		**ΔC_T (TRIB2-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**TRIB2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	19.15	0.72							
	DA	19.05	0.06							
	JP	19.56	0.26				0.19	1.00	1.14	0.87
	PK	17.46	0.13	18.81	0.19	0.00	0.27	1.00	0.19	0.13
BTC, n=5	JH	17.44	0.29							
	LB	15.89	0.30							
	MT	17.85	0.14							
	PS	18.02	0.15				0.09	3.39	3.61	3.18
	RB	16.03	0.03	17.05	0.09	-1.76	0.21	3.39	0.50	0.22
		ΔC_T (VEGFA-18S)		**ΔC_T (VEGFA-18S)**		**ΔC_T - $\Delta C_{T,cal}$**		**VEGFA rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AS1	17.36	0.19							
	DA	17.00	0.11							
	JP	18.09	0.08				0.08	1.00	1.06	0.95
	PK	16.67		17.28	0.08	0.00	0.11	1.00	0.08	0.05
BTC, n=5	JH	15.83	0.05							
	LB	15.44	0.13							
	MT	17.13	0.14							
	PS	16.11					0.05	2.62	2.71	2.53
	RB	14.95	0.04	15.89	0.05	-1.39	0.09	2.62	0.17	0.09

Table 32. Biliary brushings validation set qPCR calculations for TaqMan Array qPCR normalised to 18S rRNA.

6

		ΔC_T (GAPDH-18S)		ΔC_T (GAPDH-18S)		$\Delta\Delta C_T$ $\Delta C_T - \Delta C_{T,cal}$		GAPDH rel to Cal		
	SAMPLE	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Benign	AK	17.94	0.47							
n=4	PH	18.59	0.23							
	JB	17.34	0.81				0.25	1.00	1.19	0.84
	JS	17.48	0.34	17.84	0.25	0.00	0.36	1.00	0.25	0.18
BTC	DC	16.33	0.78							
n=5	JE	16.56	0.71							
	MJ	15.93	0.21							
	RC	18.22	0.25				0.09	1.90	2.01	1.79
	RB	17.52	0.63	16.91	0.25	-0.92	0.36	1.90	0.47	0.33

		ΔC_T (MUC4-18S)		ΔC_T (MUC4-18S)		$\Delta\Delta C_T$ $\Delta C_T - \Delta C_{T,cal}$		MUC4 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	21.15	0.67							
	PH	25.34	0.83							
	JB	23.77	0.95				0.37	1.00	1.29	0.78
	JS	30.64	0.30	25.22	0.37	0.00	0.52	1.00	0.36	0.25
BTC, n=5	DC	21.74	0.39							
	JE	21.52	0.74							
	MJ	19.74	0.22							
	RC	19.77	0.31				0.09	15.53	16.49	14.63
	RB	23.55	0.66	21.26	0.27	-3.96	0.46	15.53	4.92	2.96

		ΔC_T (MUC5AC-18S)		ΔC_T (MUC5AC-18S)		$\Delta\Delta C_T$ $\Delta C_T - \Delta C_{T,cal}$		MUC5AC rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	15.99	0.47							
	PH	20.21	0.06							
	JB	20.59	0.82				0.25	1.00	1.19	0.84
	JS	21.77	0.30	19.64	0.25	0.00	0.35	1.00	0.24	0.17
BTC, n=5	DC	17.45	0.39							
	JE	18.36	0.69							
	MJ	20.01	0.29							
	RC	14.41	0.23				0.21	5.01	5.80	4.33
	RB	16.35	0.63	17.31	0.21	-2.33	0.33	5.01	1.14	0.73

		ΔC_T (ACTB-18S)		ΔC_T (ACTB-18S)		$\Delta\Delta C_T$ $\Delta C_T - \Delta C_{T,cal}$		ACTB rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	14.31	0.47							
	PH	16.12	0.08							
	JB	14.03	0.81				0.25	1.00	1.19	0.84
	JS	14.14	0.31	14.65	0.25	0.00	0.35	1.00	0.24	0.17
BTC, n=5	DC	13.38	0.39							
	JE	13.45	0.69							

	MJ	13.52	0.27							
	RC	13.54	0.12				0.21	2.31	2.68	2.00
	RB	13.29	0.63	13.44	0.21	-1.21	0.32	2.31	0.52	0.34
						$\Delta\Delta C_T$				
		ΔC_T (CD9-18S)		ΔC_T (CD9-18S)		$\Delta C_T - \Delta C_{T,cal}$		CD9 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	20.21	0.47							
	PH	23.07	0.39							
	JB	20.26	0.83				0.27	1.00	1.20	0.83
	JS	21.56	0.32	21.28	0.27	0.00	0.38	1.00	0.26	0.19
BTC, n=5	DC	18.68	0.39							
	JE	19.85	0.69							
	MJ	20.22	0.16							
	RC	18.13	0.14				0.22	3.95	4.59	3.39
	RB	19.61	0.63	19.30	0.22	-1.98	0.35	3.95	0.95	0.60
						$\Delta\Delta C_T$				
		ΔC_T (Notch3-18S)		ΔC_T (Notch3-18S)		$\Delta C_T - \Delta C_{T,cal}$		Notch3 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	21.95	0.47							
	PH	29.35								
	JB	19.71	0.82				0.82	1.00	1.76	0.57
	JS	26.30	2.26	24.33	0.82	0.00	1.15	1.00	0.80	0.57
BTC, n=5	DC	24.33	0.45							
	JE	20.59	0.99							
	MJ	22.18	0.56							
	RC	20.11	0.63				0.30	4.72	5.79	3.84
	RB	23.23	0.84	22.09	0.30	-2.24	0.87	4.72	2.84	0.97
						$\Delta\Delta C_T$				
		ΔC_T (ASPHD1-18S)		ΔC_T (ASPHD1-18S)		$\Delta C_T - \Delta C_{T,cal}$		ASPDH1 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	23.09	0.65							
	PH	22.49	0.83							
	JB	21.04	1.05				0.39	1.00	1.31	0.77
	JS	17.82	0.41	21.11	0.39	0.00	0.55	1.00	0.38	0.27
BTC, n=5	DC	19.11	0.38							
	JE	20.45	0.78							
	MJ	19.37	0.18							
	RC	18.95	0.37				0.27	3.63	4.39	3.00
	RB	18.38	0.63	19.25	0.27	-1.86	0.47	3.63	1.19	0.69
						$\Delta\Delta C_T$				
		ΔC_T (ATP6V0A2-18S)		ΔC_T (ATP6VOA2-18S)		$\Delta C_T - \Delta C_{T,cal}$		ATP6V0A2 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	22.73	1.51							
	PH	22.69	0.74							
	JB	20.14	0.88				0.63	1.00	1.55	0.65
	JS			21.85	0.63	0.00	0.89	1.00	0.62	0.44
BTC, n=5	DC	20.47	0.47							

	JE	20.03	0.70							
	MJ	19.46	0.16							
	RC	20.36	0.13				0.26	3.14	3.75	2.62
	RB	20.71	0.63	20.21	0.26	-1.65	0.68	3.14	1.49	0.56
		ΔC_T (CEACAM1-18S)		**ΔC_T (CEACAM1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**CEACAM1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	21.70	0.60							
	PH	25.37	1.26							
	JB	22.38	0.83				0.41	1.00	1.33	0.75
	JS	22.19	0.32	22.91	0.41	0.00	0.59	1.00	0.41	0.29
BTC, n=5	DC	20.65	0.40							
	JE	21.38	0.69							
	MJ	21.47	0.29							
	RC	22.45	0.44				0.33	3.06	3.85	2.44
	RB	20.52	0.65	21.29	0.33	-1.61	0.53	3.06	1.13	0.70
		ΔC_T (CELSR1-18S)		**ΔC_T (CELSR1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**CELSR1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	24.84	0.79							
	PH		0.00							
	JB	24.59	0.92				0.32	1.00	1.24	0.80
	JS	26.21	0.37	25.21	0.32	0.00	0.45	1.00	0.31	0.22
BTC, n=5	DC	22.78	0.41							
	JE	25.41	1.10							
	MJ	23.48	0.26							
	RC	22.89	0.25				0.27	3.19	3.85	2.64
	RB	23.14	0.63	23.54	0.27	-1.67	0.42	3.19	0.92	0.60
		ΔC_T (CFDP1 - 18S)		**ΔC_T (CFDP1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**CFDP1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	21.57	0.56							
	PH	24.22	0.32							
	JB	20.33	0.80				0.27	1.00	1.21	0.83
	JS	19.85	0.32	21.49	0.27	0.00	0.38	1.00	0.26	0.19
BTC, n=5	DC	21.32	0.40							
	JE	20.84	0.69							
	MJ	19.55	0.19							
	RC	21.29	0.27				0.22	1.70	1.98	1.46
	RB	20.63	0.63	20.73	0.22	-0.77	0.35	1.70	0.41	0.25
		ΔC_T (COL17A1-18S)		**ΔC_T (COL17A1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**COL17A1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	22.04	0.48							
	PH	30.03	0.00							
	JB	22.48	1.51				0.41	1.00	1.33	0.75
	JS	22.86	0.41	24.35	0.41	0.00	0.58	1.00	0.40	0.28
BTC, n=5	DC	19.97	0.42							
	JE	21.21	1.20							

	MJ	26.31	1.46							
	RC	22.83	0.52				0.41	4.93	6.54	3.72
	RB	19.93	0.64	22.05	0.41	-2.30	0.58	4.93	1.97	1.39

		ΔC_T (COL1A1-18S)		ΔC_T (COL1A1-18S)		$\Delta C_T - \Delta C_{T,cal}$		COL1A1 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	24.23	0.47							
	PH	30.03	0.00							
	JB	17.09	0.81				0.31	1.00	1.24	0.81
	JS	27.34		24.67	0.31	0.00	0.44	1.00	0.30	0.22
BTC, n=5	DC	24.39	0.41							
	JE	21.58	0.72							
	MJ	22.43	0.22							
	RC	21.58	0.47				0.21	3.40	3.94	2.94
	RB	24.53	0.63	22.90	0.21	-1.77	0.38	3.40	0.89	0.50

		ΔC_T (COL6A3-18S)		ΔC_T (COL6A3-18S)		$\Delta C_T - \Delta C_{T,cal}$		COL6A3 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	23.95								
	PH	30.03	0.00							
	JB	14.27	0.80				0.30	1.00	1.23	0.81
	JS	23.45	0.39	22.92	0.30	0.00	0.42	1.00	0.29	0.21
BTC, n=5	DC	23.63	0.53							
	JE	21.50	0.88							
	MJ	21.44	0.58							
	RC	19.08	1.25				0.27	2.26	2.72	1.87
	RB	23.11	0.65	21.75	0.27	-1.17	0.40	2.26	0.63	0.42

		ΔC_T (CSPG4-18S)		ΔC_T (CSPG4-18S)		$\Delta C_T - \Delta C_{T,cal}$		CSPG4 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	25.59								
	PH	30.03	0.00							
	JB	21.48	1.34				0.93	1.00	1.90	0.53
	JS	23.03	2.43	25.03	0.93	0.00	1.31	1.00	0.91	0.64
BTC, n=5	DC	21.07	0.44							
	JE	30.20	0.69							
	MJ	24.24	1.40							
	RC	19.64	1.02				0.35	1.59	2.02	1.25
	RB	26.68	0.67	24.36	0.35	-0.67	0.99	1.59	1.09	0.39

		ΔC_T (HOXA10-18S)		ΔC_T (HOXA10-18S)		$\Delta C_T - \Delta C_{T,cal}$		HOXA10 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	25.78	2.32							
	PH	30.03	0.00							
	JB	23.54					0.78	1.00	1.72	0.58
	JS	26.68	0.35	26.51	0.78	0.00	1.10	1.00	0.77	0.54
BTC, n=5	DC	18.83	0.44							

	JE	23.18	0.94							
	MJ	26.87	2.49							
	RC	17.96	1.02				0.55	35.84	52.60	24.42
	RB	19.87	0.63	21.34	0.55	-5.16	0.96	35.84	23.78	13.75

		ΔC_T (HOXB6-18S)		ΔC_T (HOXB6-18S)		$\Delta C_T - \Delta C_{T,cal}$		HOXB6 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	19.40	0.49							
	PH	27.09								
	JB	21.15	0.83				0.48	1.00	1.40	0.72
	JS	24.27		22.98	0.48	0.00	0.68	1.00	0.47	0.33
BTC, n=5	DC	18.67	0.44							
	JE	20.25	0.69							
	MJ	22.74	0.28							
	RC	16.18	0.36				0.21	10.19	11.81	8.79
	RB	20.30	0.63	19.63	0.21	-3.35	0.53	10.19	3.72	1.50

		ΔC_T (ITGB8-18S)		ΔC_T (ITGB8-18S)		$\Delta C_T - \Delta C_{T,cal}$		ITGB8 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	20.84	0.46							
	PH	20.68	3.69							
	JB	20.43	0.92				0.96	1.00	1.95	0.51
	JS	18.96	0.38	20.23	0.96	0.00	1.36	1.00	0.94	0.67
BTC, n=5	DC	18.73	0.52							
	JE	20.46	0.70							
	MJ	17.69	1.29							
	RC	18.79	0.86				0.81	2.60	4.56	1.48
	RB	18.58	0.63	18.85	0.81	-1.38	1.26	2.60	2.27	1.46

		ΔC_T (ITIH5-18S)		ΔC_T (ITIH5-18S)		$\Delta C_T - \Delta C_{T,cal}$		ITIH5 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	25.12								
	PH	24.90								
	JB	21.59	1.37				0.83	1.00	1.78	0.56
	JS	20.54	0.94	23.04	0.83	0.00	1.17	1.00	0.81	0.58
BTC, n=5	DC	23.70	0.60							
	JE	22.46								
	MJ	21.82	0.38							
	RC	22.46	0.56				0.44	1.45	1.98	1.07
	RB	22.05	1.63	22.49	0.44	-0.54	0.94	1.45	0.95	0.45

		ΔC_T (KRAS-18S)		ΔC_T (KRAS-18S)		$\Delta C_T - \Delta C_{T,cal}$		KRAS rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	19.97	0.47							
	PH	20.68	0.20							
	JB	18.09	0.85				0.27	1.00	1.20	0.83
	JS	18.67	0.38	19.35	0.27	0.00	0.38	1.00	0.26	0.18

BTC, n=5	DC	18.62	0.55							
	JE	18.54	0.74							
	MJ	18.06	0.17							
	RC	19.04	0.52				0.23	1.46	1.71	1.25
	RB	19.76	0.63	18.80	0.23	-0.55	0.35	1.46	0.35	0.23

		ΔC_T (LAMC2-18S)		**ΔC_T (LAMC2-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**LAMC2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	20.84	0.47							
	PH	23.56	0.08							
	JB	22.06	0.82				0.26	1.00	1.20	0.84
	JS	21.78	0.42	22.06	0.26	0.00	0.37	1.00	0.25	0.18
BTC, n=5	DC	19.66	0.39							
	JE	21.15	0.73							
	MJ	21.31	0.22							
	RC	19.61	0.13				0.21	3.27	3.78	2.82
	RB	20.05	0.63	20.35	0.21	-1.71	0.33	3.27	0.76	0.48

		ΔC_T (LEF1-18S)		**ΔC_T (LEF1-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**LEF1 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	24.94								
	PH	30.03	0.00							
	JB	21.73	0.94				0.82	1.00	1.77	0.57
	JS	27.72	2.28	26.10	0.82	0.00	1.16	1.00	0.80	0.57
BTC, n=5	DC	24.19	0.57							
	JE	21.71	0.91							
	MJ	23.99	0.54							
	RC	25.12	0.16				0.27	4.94	5.96	4.09
	RB	23.99	0.63	23.80	0.27	-2.30	0.86	4.94	2.96	0.93

		ΔC_T (LIF-18S)		**ΔC_T (LIF-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**LIF rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	20.96	0.57							
	PH	22.97	0.15							
	JB	21.02	0.80				0.26	1.00	1.20	0.83
	JS	20.06	0.33	21.25	0.26	0.00	0.37	1.00	0.26	0.18
BTC, n=5	DC	19.79	0.41							
	JE	20.51	0.70							
	MJ	20.17	0.20							
	RC	20.16	0.46				0.21	1.97	2.28	1.70
	RB	20.75	0.64	20.27	0.21	-0.98	0.34	1.97	0.46	0.29

		ΔC_T (LUM-18S)		**ΔC_T (LUM-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**LUM rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	24.90	1.15							
	PH	30.03	0.00							
	JB	17.37	1.05				0.52	1.00	1.43	0.70

	JS	23.19		23.87	0.52	0.00	0.73	1.00	0.51	0.36
BTC, n=5	DC	23.99	0.70							
	JE	24.00	0.88							
	MJ	23.71	0.51							
	RC	21.61	0.38				0.30	1.07	1.32	0.87
	RB	25.54	0.87	23.77	0.30	-0.10	0.60	1.07	0.45	0.22

		ΔC_T (MAML2-18S)		ΔC_T (MAML2-18S)		$\Delta C_T - \Delta C_{T,cal}$		MAML2 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	23.34	0.47							
	PH	24.93								
	JB	21.72	1.05				0.40	1.00	1.32	0.76
	JS	22.86	0.39	23.21	0.40	0.00	0.57	1.00	0.40	0.28
BTC, n=5	DC	22.31	0.38							
	JE	23.58	0.93							
	MJ	22.75	0.32							
	RC	22.97	0.43				0.25	1.39	1.66	1.18
	RB	22.05	0.64	22.73	0.25	-0.48	0.47	1.39	0.46	0.24

		ΔC_T (MAML3-18S)		ΔC_T (MAML3-18S)		$\Delta C_T - \Delta C_{T,cal}$		MAML3 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	21.88	0.48							
	PH	24.00	0.95							
	JB	21.41	0.85				0.35	1.00	1.27	0.79
	JS	21.03	0.30	22.08	0.35	0.00	0.49	1.00	0.34	0.24
BTC, n=5	DC	21.21	0.44							
	JE	20.73	0.74							
	MJ	20.35	0.23							
	RC	17.43	0.37				0.29	3.74	4.57	3.06
	RB	21.18	0.63	20.18	0.29	-1.90	0.45	3.74	1.17	0.75

		ΔC_T (MAPK1-18S)		ΔC_T (MAPK1-18S)		$\Delta C_T - \Delta C_{T,cal}$		MAPK1 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	21.79	0.50							
	PH	22.74	0.06							
	JB	20.37	0.84				0.26	1.00	1.20	0.83
	JS	20.57	0.37	21.37	0.26	0.00	0.37	1.00	0.26	0.18
BTC, n=5	DC	20.60	0.39							
	JE	20.05	0.71							
	MJ	19.74	0.17							
	RC	21.17	0.16				0.21	2.15	2.49	1.86
	RB	19.76	0.63	20.26	0.21	-1.11	0.33	2.15	0.50	0.31

		ΔC_T (MCM4-18S)		ΔC_T (MCM4-18S)		$\Delta C_T - \Delta C_{T,cal}$		MCM4 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	23.86	0.58							

	PH	24.46	0.90							
	JB	21.40	1.54				0.48	1.00	1.39	0.72
	JS	24.02	0.35	23.43	0.48	0.00	0.68	1.00	0.47	0.33
BTC, n=5	DC	22.68	0.54							
	JE	21.74	0.79							
	MJ	22.15	0.67							
	RC	22.99	1.15				0.32	2.49	3.11	2.00
	RB	21.04	0.63	22.12	0.32	-1.32	0.58	2.49	0.99	0.56

		ΔC_T (MMP2-18S)		**ΔC_T (MMP2-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**MMP2 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	30.38								
	PH	30.03								
	JB	15.35	1.08				1.08	1.00	2.12	0.47
	JS			25.25	1.08	0.00	1.53	1.00	1.06	0.75
BTC, n=5	DC	25.71	0.50							
	JE	24.48	2.20							
	MJ	23.62	0.23							
	RC	22.24	0.25				0.49	2.34	3.28	1.66
	RB	24.10	0.92	24.03	0.49	-1.22	1.19	2.34	1.93	0.79

		ΔC_T (MXRA5-18S)		**ΔC_T (MXRA5-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**MXRA5 rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	22.66								
	PH	26.25	0.53							
	JB	17.71	0.81				0.34	1.00	1.26	0.79
	JS	30.64	0.30	24.31	0.34	0.00	0.48	1.00	0.33	0.23
BTC, n=5	DC	23.73	0.57							
	JE	21.15	0.70							
	MJ	22.92	0.71							
	RC	21.42	0.41				0.28	3.55	4.32	2.91
	RB	23.21	0.65	22.49	0.28	-1.83	0.44	3.55	1.08	0.70

		ΔC_T (MYC-18S)		**ΔC_T (MYC-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**MYC rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	22.83	0.76							
	PH	25.77	0.03							
	JB	20.95	0.81				0.29	1.00	1.22	0.82
	JS	25.39	0.32	23.74	0.29	0.00	0.41	1.00	0.28	0.20
BTC, n=5	DC	19.44	0.38							
	JE	20.08	0.70							
	MJ	20.65	0.17							
	RC	22.09	0.29				0.21	11.52	13.30	9.99
	RB	18.80	0.63	20.21	0.21	-3.53	0.36	11.52	2.84	1.65

		ΔC_T (NCSTN-18S)		**ΔC_T (NCSTN-18S)**		**$\Delta C_T - \Delta C_{T,cal}$**		**NCSTN rel to Cal**		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	

Normal, n=4	AK	21.19	0.71							
	PH	22.66	0.20							
	JB	19.48	0.81				0.28	1.00	1.22	0.82
	JS	19.50	0.30	20.71	0.28	0.00	0.40	1.00	0.28	0.20
BTC, n=5	DC	20.00	0.39							
	JE	20.23	0.74							
	MJ	19.01	0.17							
	RC	19.66	0.22				0.22	1.89	2.20	1.63
	RB	20.03	0.63	19.79	0.22	-0.92	0.36	1.89	0.47	0.28

		ΔC_T (PEAR1-18S)		ΔC_T (PEAR1-18S)		$\Delta C_T - \Delta C_{T,cal}$		PEAR1 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	24.48	0.99							
	PH	30.03	0.00							
	JB	23.04	0.82				0.43	1.00	1.35	0.74
	JS	25.58		25.78	0.43	0.00	0.60	1.00	0.42	0.30
BTC, n=5	DC	22.47	0.67							
	JE	24.77	0.72							
	MJ	28.34								
	RC	22.71	0.19				0.29	2.67	3.27	2.18
	RB	23.55	0.64	24.37	0.29	-1.42	0.52	2.67	0.96	0.54

		ΔC_T (PNMA2-18S)		ΔC_T (PNMA2-18S)		$\Delta C_T - \Delta C_{T,cal}$		PNMA2 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	25.06	0.47							
	PH	30.03	0.00							
	JB	26.47	0.86				0.25	1.00	1.19	0.84
	JS	30.64	0.30	28.05	0.25	0.00	0.36	1.00	0.25	0.18
BTC, n=5	DC	20.61	0.42							
	JE	23.40	1.11							
	MJ	23.49	0.18							
	RC	17.29	0.12				0.27	112.33	135.72	92.97
	RB	21.39	0.65	21.24	0.27	-6.81	0.37	112.33	29.07	21.25

		ΔC_T (POU5F1-18S)		ΔC_T (POU5F1-18S)		$\Delta C_T - \Delta C_{T,cal}$		POU5F1 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	23.79	0.51							
	PH	24.66	0.93							
	JB	21.85	0.85				0.35	1.00	1.28	0.78
	JS	22.82	0.38	23.28	0.35	0.00	0.50	1.00	0.35	0.24
BTC, n=5	DC	20.48	0.38							
	JE	23.05	0.85							
	MJ	23.02	0.45							
	RC	17.22	0.16				0.31	5.54	6.85	4.48
	RB	20.27	0.63	20.81	0.31	-2.47	0.47	5.54	1.79	1.18

		ΔC_T (PRKCB1-18S)		ΔC_T (PRKCB1-18S)		$\Delta C_T - \Delta C_{T,cal}$		PRKCB1 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	21.92	0.47							
	PH	25.57								
	JB	19.81	0.81				0.34	1.00	1.27	0.79
	JS	23.64	0.43	22.73	0.34	0.00	0.48	1.00	0.34	0.24
BTC, n=5	DC	20.13	0.42							
	JE	19.90	0.87							
	MJ	21.98	0.24							
	RC	20.36	0.53				0.24	3.13	3.69	2.66
	RB	23.06	0.64	21.09	0.24	-1.65	0.42	3.13	0.90	0.51
		ΔC_T (PTK2-18S)		ΔC_T (PTK2-18S)		$\Delta C_T - \Delta C_{T,cal}$		PTK2 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	20.17	0.48							
	PH	22.74	0.34							
	JB	20.88	0.84				0.27	1.00	1.20	0.83
	JS	19.55	0.30	20.83	0.27	0.00	0.38	1.00	0.26	0.18
BTC, n=5	DC	20.21	0.39							
	JE	20.21	0.76							
	MJ	19.48	0.21							
	RC	20.35	0.18				0.23	1.84	2.16	1.58
	RB	19.52	0.64	19.95	0.23	-0.88	0.35	1.84	0.45	0.29
		ΔC_T (PVT1 -18S)		ΔC_T (PVT1-18S)		$\Delta C_T - \Delta C_{T,cal}$		PVT1 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	26.04	0.89							
	PH	30.03	0.00							
	JB	23.60	0.96				0.34	1.00	1.26	0.79
	JS	30.64	0.30	27.57	0.34	0.00	0.48	1.00	0.33	0.23
BTC, n=5	DC	22.20	0.38							
	JE	25.20	1.13							
	MJ	26.44	1.12							
	RC	23.83	0.18				0.35	12.55	16.01	9.84
	RB	21.95	0.63	23.92	0.35	-3.65	0.49	12.55	4.23	3.06
		ΔC_T (RAB27A-18S)		ΔC_T (RAB27A-18S)		$\Delta C_T - \Delta C_{T,cal}$		RAB27A rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	21.22	0.49							
	PH	23.62	0.16							
	JB	21.47	0.80				0.25	1.00	1.19	0.84
	JS	22.76	0.30	22.27	0.25	0.00	0.35	1.00	0.24	0.17
BTC, n=5	DC	20.97	0.41							
	JE	21.80	0.76							
	MJ	20.90	0.27							
	RC	20.57	0.16				0.22	2.73	3.19	2.34
	RB	19.84	0.63	20.82	0.22	-1.45	0.33	2.73	0.63	0.42

		ΔC_T (RAD51-18S)		ΔC_T (RAD51-18S)		$\Delta C_T - \Delta C_{T,cal}$		RAD51 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	25.59	0.48							
	PH	30.03	0.00							
	JB	23.80	0.88				0.33	1.00	1.26	0.79
	JS	25.65		26.27	0.33	0.00	0.47	1.00	0.33	0.23
BTC, n=5	DC	23.75	0.89							
	JE	23.09	0.74							
	MJ	26.54	0.47							
	RC	25.10	0.63				0.28	4.19	5.11	3.44
	RB	22.52	0.69	24.20	0.28	-2.07	0.44	4.19	1.27	0.83

		ΔC_T (SERPINA3-18S)		ΔC_T (SERPINA3-18S)		$\Delta C_T - \Delta C_{T,cal}$		SERPINA3 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	19.44	0.61							
	PH	22.98	1.00							
	JB	18.28	1.30				0.44	1.00	1.36	0.74
	JS	22.32	0.30	20.76	0.44	0.00	0.63	1.00	0.43	0.31
BTC, n=5	DC	17.73	0.39							
	JE	18.77	0.90							
	MJ	17.78	0.29							
	RC	17.68	0.72				0.31	4.40	5.46	3.54
	RB	21.14	0.63	18.62	0.31	-2.14	0.54	4.40	1.65	0.95

		ΔC_T (SPARC-18S)		ΔC_T (SPARC-18S)		$\Delta C_T - \Delta C_{T,cal}$		SPARC rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	23.84	0.47							
	PH	26.66	1.83							
	JB	16.27	1.12				0.78	1.00	1.72	0.58
	JS	25.94	2.23	23.18	0.78	0.00	1.11	1.00	0.77	0.54
BTC, n=5	DC	22.52	0.45							
	JE	19.68	0.69							
	MJ	21.67	0.37							
	RC	20.97	0.50				0.43	3.14	4.23	2.33
	RB	22.80	0.69	21.53	0.43	-1.65	0.89	3.14	1.94	0.94

		ΔC_T (STAT1-18S)		ΔC_T (STAT1-18S)		$\Delta C_T - \Delta C_{T,cal}$		STAT1 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	18.49	0.51							
	PH	20.74	0.03							
	JB	17.15	0.94				0.28	1.00	1.21	0.82
	JS	17.32	0.33	18.42	0.28	0.00	0.40	1.00	0.27	0.19
BTC, n=5	DC	17.14	0.49							
	JE	18.02	0.80							
	MJ	15.16	0.20							
	RC	17.92	0.38				0.24	2.79	3.28	2.36
	RB	16.49	0.70	16.95	0.24	-1.48	0.37	2.79	0.71	0.46

		ΔC_T (TGFBI-18S)		ΔC_T (TGFBI-18S)		$\Delta C_T - \Delta C_{T,cal}$		TGFBI rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	22.76	0.60							
	PH	23.93								
	JB	18.08	0.83				0.46	1.00	1.38	0.73
	JS	24.72	0.93	22.37	0.46	0.00	0.65	1.00	0.45	0.32
BTC, n=5	DC	21.53	0.62							
	JE	19.64	0.71							
	MJ	21.22	0.16							
	RC	20.45	0.32				0.23	2.49	2.93	2.13
	RB	22.44	0.64	21.05	0.23	-1.32	0.52	2.49	0.89	0.40

		ΔC_T (TM4SF18-18S)		ΔC_T (TM4SF18-18S)		$\Delta C_T - \Delta C_{T,cal}$		TM4SF18 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	23.57								
	PH	26.63								
	JB	20.13	0.86				0.50	1.00	1.42	0.71
	JS	22.90	0.51	23.31	0.50	0.00	0.71	1.00	0.49	0.35
BTC, n=5	DC	22.95	0.45							
	JE	21.37	1.11							
	MJ	20.28	0.59							
	RC	20.21	0.59				0.29	5.26	6.45	4.29
	RB	19.75	0.63	20.91	0.29	-2.39	0.58	5.26	2.12	1.07

		ΔC_T (TRIB2-18S)		ΔC_T (TRIB2-18S)		$\Delta C_T - \Delta C_{T,cal}$		TRIB2 rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	20.92	0.50							
	PH	23.36	1.01							
	JB	20.88	0.87				0.36	1.00	1.29	0.78
	JS	21.51	0.30	21.67	0.36	0.00	0.52	1.00	0.36	0.25
BTC, n=5	DC	19.46	0.50							
	JE	19.36	0.71							
	MJ	20.36	0.16							
	RC	20.97	0.41				0.30	3.12	3.83	2.54
	RB	20.00	0.63	20.03	0.30	-1.64	0.47	3.12	1.02	0.64

		ΔC_T (VEGFA-18S)		ΔC_T (VEGFA-18S)		$\Delta C_T - \Delta C_{T,cal}$		VEGFA rel to Cal		
	SAMPLES	ΔCT	STDEV	AVERAGE	ERROR	Average	ERROR		ERROR	
Normal, n=4	AK	20.22	0.47							
	PH	22.24	0.10							
	JB	19.40	0.86				0.33	1.00	1.26	0.80
	JS	19.53		20.35	0.33	0.00	0.46	1.00	0.32	0.23
BTC, n=5	DC	18.70	0.40							
	JE	18.75	0.70							
	MJ	18.81								

RC	18.75	0.17				0.26	2.92	3.48	2.44
RB	18.99	0.63	18.80	0.26	-1.54	0.42	2.92	0.84	0.52

9.1.2 Abstracts related to the thesis.

A single centre experience of dominant biliary strictures and cholangiocarcinoma in primary sclerosing cholangitis

M.H. Chapman[1,2], J. Witmann[1], G.J. Webster[1,2] A.R. Hatfield[2] & S.P. Pereira[1,2]

[1]The Institute of Hepatology, University College London (UCL) and [2]Dept Gastroenterology, UCL Hospitals NHS Foundation Trust, London, UK.

Background

Dominant biliary strictures occur commonly in patients with primary sclerosing cholangitis (PSC) (fig1), who have a high lifetime risk (up to 20%) of developing cholangiocarcinoma[1]. The optimal management and natural history of dominant strictures remains unclear, with uncontrolled reports suggesting that endoscopic intervention improves outcome[2,3] although some controversy remains[4].

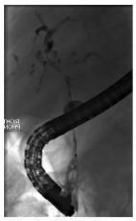


Fig 1. ERCP of dominant stricture

Aims

- To describe the long term follow up and natural history of PSC and PSC related cholangiocarcinoma in our centre.
- To assess the safety and efficacy of endoscopic retrograde cholangiopancreatography (ERCP) for dominant strictures in PSC

Methods

Review of data collected retrospectively on 112 patients undergoing one or more ERCPs for PSC at University College London Hospitals between 1982 and 2004.

Patient characteristics

	No Dominant Stricture	Dominant Stricture	
Number of patients	46	66	
Mean follow up in years (range)	6.8 (0-19)	4.8 (0-19.8)	P=0.03
Mean bilirubin (µmol/L)	20	71	P<0.01
Mean ALP (IU/L)	504	606	P=0.17
Mean ALT (IU/L)	97	102	P=0.399
UDCA	13/46 (28%)	32/66 (50%)	p=0.10
Median number of ERCP interventions	1 (0-7)	4 (0-32)	P<0.01
Median number of ERCP stents	0 (0-2)	3 (0-16)	P<0.01
Cholangiocarcinoma (%)	0 /46 (0%)	13/66 (20%)	p<0.001
Deceased	7/46 (15%)	33/66 (50%)	p=0.02
Mean survival (years)	19.4	12.8	P=0.02

Table 1. Patient characteristics

Results: Dominant strictures

- 66 patients with dominant biliary strictures had a median of 4 (range 0-32) interventions compared to 1 (0-7) in the 46 without dominant strictures (p<0.001).
- The presence of choledocholithiasis was an indication for ERCP in 18 (27%) of those with dominant stricture and 10 (22%) without.
- Major complications were 2 cases of perforated bile duct (1% of all ERCP procedures); neither required surgical treatment.
- There were no procedure-related deaths.
- 2 patients underwent liver transplantation; 1 in each group.
- There were 29 deaths (44%) in those with dominant strictures and 7 (15%) in those without.
- Survival was significantly worse in those with dominant strictures, even when patients with cholangiocarcinoma were excluded (table 1 and fig 3).
- The cause of death in those with and without dominant strictures were cholangiocarcinoma (44% & 0%), liver failure (38% & 43%) and others (18% & 57%) respectively.

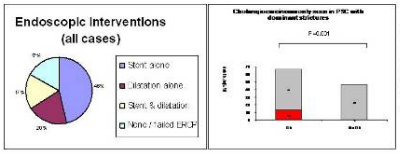


Fig 2. A) Frequency of endoscopic interventions and B) rates of cholangiocarcinoma with and without dominant strictures

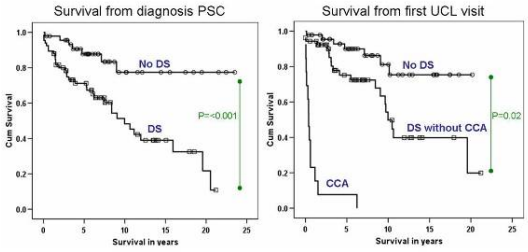


Fig 3. Kaplan Meier survival curves for patients with PSC at UCLH. (CCA; cholangiocarcinoma)

Results: Cholangiocarcinoma

- Cholangiocarcinoma was only seen in patients with dominant strictures (frequency 20%) (fig 2B).
- Jaundice was the presenting feature in 11 (85%).
- Mean CA19.9 at diagnosis of cholangiocarcinoma was 14,971 IU/L (9-84,000) compared to 77 IU/L (1-1015) in the benign group (p=0.16).
- Patients developed cholangiocarcinoma a median of 3.6 years (0-21 years) after the diagnosis of PSC.
- Six (46%) presented within 4 months of diagnosis of PSC (Fig 4A).
- The incidence of cholangiocarcinoma was 0.92% per year (excluding those presenting within 1 year of diagnosis PSC).
- Median time from diagnosis to death was 5 months (1-23 months).
- Mean scores for liver disease models at diagnosis of cholangiocarcinoma were significantly higher than in benign disease (fig 4B).

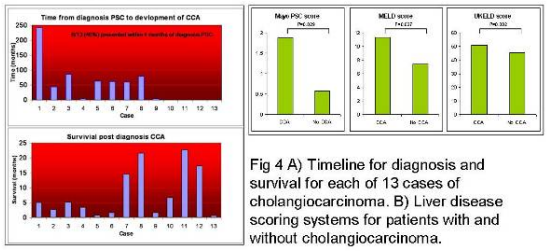


Fig 4 A) Timeline for diagnosis and survival for each of 13 cases of cholangiocarcinoma. B) Liver disease scoring systems for patients with and without cholangiocarcinoma.

Conclusions

The overall survival in our cohort, with a mean follow up of more than 5 years, is similar to that published in other series. Repeated endoscopic therapy in patients with PSC appears to be safe. However, our data highlight the poor prognosis in the subgroup with dominant strictures even after exclusion of those who developed cholangiocarcinoma. In our series, cholangiocarcinoma only occurred in patients with dominant bile duct strictures with 6/13 (46%) presenting in the first year with an annual incidence thereafter of approximately 1%. Further prospective trials in the management of this difficult disease are required.

References

Bergquist et al. Hepatic and extrahepatic malignancies in PSC. J Hep 36, 2002.
Baluyut et al. Impact of endoscopic therapy on survival of patients with PSC. GI End 53,2001.
Ponsioen et al. Four years experience with short term stenting in PSC. Am J Gastro, 94, 1999.
Bjornsson et al. Dominant strictures in patients with PSC. Am J Gastro 2004

Figure 72 (Abstract 1). Copy of abstract presented at the Bile Acid Biology and Therapeutics meeting, Amsterdam, June 2008 outlining our local experience of dominant strictures and cholangiocarcinoma in PSC.

High success rate for tissue diagnosis in 198 patients with biliary tract cancer.

Mills R[1], Chapman M.H.[1,2], Sandanayake N.S.[1,2], & Pereira S.P.[1,2]

[1]The Institute of Hepatology, UCL Medical School and [2]Department of Gastroenterology, University College London Hospitals NHS Foundation Trust, London, UK.

UCL

Introduction

•The diagnosis of biliary tract cancer (BTC: cholangiocarcinoma and gallbladder cancer) is challenging because of a lack of reliable tumour markers, radiological similarities with benign hepatobiliary disease, and tumour morphology and location limiting access to diagnostic tissue (*Figure 1*).

•Current guidelines state that confirmatory histology and/or cytology should be obtained if possible, as definitive diagnosis guides appropriate treatment (Khan et al, Gut 2002).

•High rates of benign disease (up to 17%) have been described in patients undergoing surgical resection for presumed BTC (Erdogan et al, B J Surg 2008)

Aim

•To assess the utility and sensitivity of techniques used for cytological/histological confirmation of BTC.

Methods

•Patients identified using prospectively collected databases and notes of the HPB and cancer MDM's between January 2003 and December 2008.

•All cytology and tissue biopsy data for each patient were recorded, including number and route of diagnostic procedures and result of each attempt.

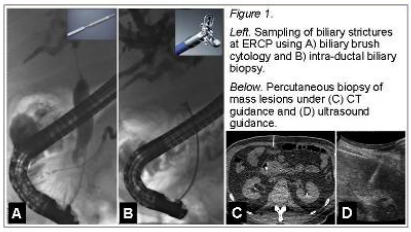


Figure 1.
Left. Sampling of biliary strictures at ERCP using A) biliary brush cytology and B) intra-ductal biliary biopsy.
Below. Percutaneous biopsy of mass lesions under (C) CT guidance and (D) ultrasound guidance.

Results: Patient characteristics

•198 patients received a final clinical diagnosis of BTC (173 cholangiocarcinoma, 25 gallbladder cancer).

•Male:female ratio = 1.1:1

•Mean age 70 years (range 25-105)

•Median serum CA19-9 at diagnosis was 709 U/ml (normal 0-47 U/ml, range <1 to 145,528).

•Median serum bilirubin level was 45µmol/L (normal 5.1 – 17 µmol/L, range 3 to 822).

•52 patients (26%) had a tissue diagnosis prior to referral

Results: Tissue diagnosis

•Of the 146 patients without a prior tissue diagnosis, 145 (99%) had attempted tissue sampling (median 1 procedure, range 1-6) at our hospital (UCLH).

•Out of a total of 286 tissue samples, 145 were taken for cytology (biliary brushings n=96; bile n=1; pleural and ascitic fluid n=38; endoscopic ultrasound-guided fine needle aspiration (FNA) n=10), of which 37 (26%) were positive for cancer.

•141 biopsies (percutaneous n=105, endobiliary n=36) were taken for histology, of which 91 (65%) were positive for cancer.

•33 samples (11%) were reported as suspicious for malignancy and a further 15 (5%) had suboptimal specimens unsuitable for pathological diagnosis.

•In total, 123/146 (84%) reached pathological confirmation of cancer.

•A clinical diagnosis of BTC, based on multidisciplinary review and later confirmed by evidence of disease progression, was made in 23 (16%), including 7 (5%) with suspicious pathology.

Table 1, Sensitivity of various tissue sampling methods used in reaching a diagnosis of BTC

Method	n	Confirmation of cancer (%)	Suspicious for cancer	No evidence of cancer	Sub-optimal sample
Biliary brush cytology	96	30 (31%)	21 (22%)	39 (41%)	6 (6%)
Bile cytology	1	0 (0%)	0 (0%)	0 (0%)	1 (100%)
Ascites cytology	3	6 (19%)	2 (7%)	23 (74%)	0 (0%)
Pleural and other fluid cytology	7	1 (14%)	0 (0%)	6 (86%)	0 (0%)
Intra-ductal biliary biopsy	36	16 (44%)	5 (14%)	10 (28%)	5 (14%)
EUS FNA and/or biopsy	10	5 (50%)	1 (10%)	2 (20%)	2 (20%)
Percutaneous biopsy	105	75 (71%)	4 (4%)	25 (24%)	1 (1%)
All procedures	286	133 (47%)	33 (11%)	105 (37%)	15 (5%)

Pathological features suspicious for cancer

•Of the 145 patients sampled at UCLH, 29 had one or more samples reported as suspicious for cancer (see table).

•Of these, 22 had later confirmatory tissue diagnosis, usually obtained by percutaneous biopsy (n=14) or surgical resection (n=2).

•A further 3 patients were confirmed with repeat biliary brush cytology and 3 with endobiliary biopsies, 2 of which were taken at the same time as the brush cytology reported as suspicious for cancer.

•In the 7 patients without confirmatory tissue diagnosis but evidence of disease progression, one had end stage disease considered unsuitable for further invasive tests, but the other 6 patients had a median of 4 (range 2-6) negative diagnostic procedures.

Conclusions

•In this large single-centre series, a pathological diagnosis of BTC was achieved in 84% of patients, exceeding best published data (Witzigmann et al, Ann Surg 2006).

•The sensitivity of biliary cytology in routine clinical practice is low, but a pathological diagnosis can be obtained with a combination of other endoscopic and percutaneous approaches, allowing alternative diagnoses to be excluded and appropriate treatment given.

Figure 73 (Abstract 2). Copy of abstract poster presented at the BASL annual meeting 2009 outlining local success rates in reaching a tissue diagnosis in BTC using currently available diagnostic modalities.

CIRCULATING CYFRA 21-1 IS A POTENTIAL MARKER FOR CHOLANGIOCARCINOMA IN PRIMARY SCLEROSING CHOLANGITIS

R Tidswell[1], M.H. Chapman[1,3], N. Sandanayake[1,3], F Andreola[1], G.J. Webster[1,3], J.S. Dooley[2] & S.P. Pereira[1,3]

[1]The Institute of Hepatology and [2]Centre for Hepatology (Royal Free Campus), University College London (UCL) and [3]Dept. of Gastroenterology, UCL Hospitals NHS Foundation Trust, London, UK.

Background

Patients with primary sclerosing cholangitis (PSC) carry a high lifetime risk of developing cholangiocarcinoma (CC)[1], which has a poor prognosis and is mostly found in patients with dominant strictures (Figs 1 & 2). The most commonly used serum marker, CA19-9, has limited accuracy for diagnosis or screening. Cytokeratin 19 is an epithelial cell marker expressed by cholangiocytes. One of its fragments, CYFRA21-1, has been reported to be a useful diagnostic and prognostic marker of intrahepatic CC in a Japanese population[2] and can also be detected in the sera of patients with other adenocarcinomas (lung, breast, bladder) but not in healthy individuals[3,4].

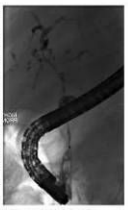


Fig 1. ERCP of dominant stricture

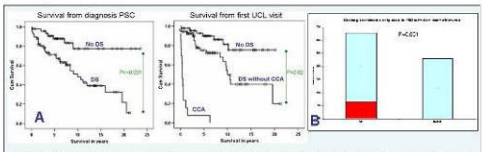


Fig 2. A) Kaplan Meier survival curves for patients with PSC at University College Hospital.
B) Rates of CC in patients with PSC with and without dominant strictures seen at UCH

Aims

To assess the role of CYFRA21-1 as a marker of CC in patients with and without PSC.

Methods

•Following local ethical approval, blood samples were collected prospectively from patients with PSC (n=19), PSC related CC (n= 6) and sporadic biliary tract cancer (BTC) (CC [n=56], gall bladder cancer, [n=4]).

•Serum CYFRA 21-1 levels were measured in duplicate by ELISA (DRG Instruments, Marburg, Germany).

•Diagnostic utility was calculated by comparing CYFRA 21-1 levels in patients with PSC to those with PSC related CC and sporadic BTC.

Patient characteristics

Table 1. Patient characteristics, serum bilirubin and CA19-9 levels.

	Benign disease n=39	PSC n=19	PSC/CC n=6	Sporadic BTC n=60
Age (years)	53 (25-80)	50 (20-75)	60 (25-73)	68 (34-91)
Gender (M:F)	20:19	10:9	4:2	30:30
Bilirubin (mmol/L)	14 (5-335)	17 (7-341)	81 (28-375)	43 (8-676)
CA19.9 (U/ml)	10.5 (0-686)	13.5 (0-3145)	473 (129- 4,139)	316 (0- 145,528)

Results: Circulating levels and ROC curves

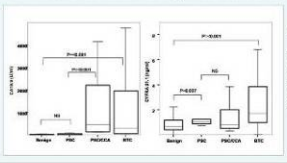


Fig 3. Circulating levels of CA19-9 and CYFRA 21-1 in each of the patient groups.

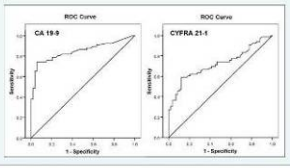


Fig 4. Receiver operator characteristic (ROC) curves for CA19-9 and CYFRA 21-1. Area under the curve is 84% (77-91%) for CA19-9 and 73% (64-82%) for CYFRA21-1.

Results: CYFRA21-1 as a marker of BTC

•CYFRA21-1 (3ng/ml), with or without CA19-9, achieved specificities of 96% to 100% for the diagnosis of BTC (Table 2).

•CYFRA21-1 is a marker of tumour stage (Figure 5).

•Elevated CYFRA21-1 is an indicator of poor prognosis (Figure 6 A and B).

Table 2. Diagnostic sensitivities and specificities for CYFRA21-1 and CYFRA21-1 in combination with CA19-9 using different diagnostic cutoffs.

	CA 19-9 >37	CA 19-9 >129	CYFRA >1.5	CYFRA >3.0	CA19-9 >37 + CYFRA >1.5	CA19-9 >129 + CYFRA >3.0
Sensitivity	79	74	58	30	45	27
Specificity	78	95	88	97	96	100
PPV	81	94	84	91	94	100
NPV	75	76	64	55	59	53

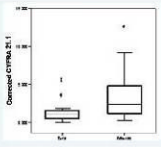
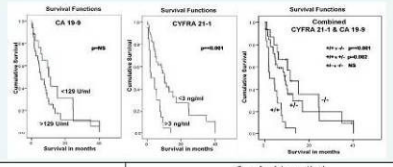


Fig 5. CYFRA 21-1 is a marker of tumour stage in patients with BTC. Patients with stage III/IV disease had median CYFRA21-1 levels of 2.42ng/ml (0.49-35) compared to 1.03ng/ml (0.32-5.59) in stage I/II disease (Mann Whitney U test p=0.001).

A



B

		Survival (months)			
		1 year	2 years	3 years	Median
CYFRA 21-1	>3 ng/ml (n=21)	3%	0	0	2
	<3 ng/ml (n=45)	42%	22%	10%	10

Fig 6. A) Kaplan-Meier plots survival and B) percentage survival for patients with BTC and CYFRA21-1 levels greater and less than 3ng/ml.

Results: CYFRA21-1 as a marker of BTC in PSC

•In the subgroup with PSC and PSC related CC, the sensitivity and specificity of CYFRA 21-1 (3 ng/ml) for the diagnosis of CC were 17% and 95%, respectively.

•The sensitivity and specificity for combined elevation with CA19-9 was 33% and 95% for the lower cutoffs (CYFRA 21-1 >1.5ng/ml and CA19-9 >37IU/ml) and 16% and 100% for the higher cutoffs (>3ng/ml and >129IU/ml).

•Results in Table 3 compare patients with benign PSC against all BTC.

•Numbers were insufficient to assess CYFRA 21-1 as a prognostic marker in PSC related CC.

Table 3. Diagnostic utility of CYFRA 21-1 and CA 19-9 for patients with BTC. Data shown compared those with benign PSC against all patients with BTC.

	CA 19-9 >37U/ml	CA 19-9 >129U/ml	CYFRA 21.1 >1.5 ng/ml	CYFRA 21-1 >3.0 ng/ml	CYFRA >1.5 + CA19-9>37	CYFRA >3.0 + CA19-9>129
Sensitivity	79	73	56	30	51	27
Specificity	79	95	68	95	95	100
PPV	40	71	24	50	63	100
NPV	95	95	90	89	92	89

Conclusions

CYFRA 21-1 is an accurate diagnostic and prognostic marker in patients with sporadic BTC and may be a useful screening tool for CC in patients with PSC. Combining CYFRA 21-1 with CA19-9 further improves diagnostic accuracy, with a specificity similar to that of biliary brush cytology (96-100%). These initial data warrant validation in multicentre studies.

References

1. Bergquist et al. Hepatic and extrahepatic malignancies in PSC. J Hep 36, 2002.
2. Uenishi et al. Serum Cytokeratin 19 Fragment (CYFRA21-1) as a Prognostic Factor in Intrahepatic Cholangiocarcinoma. Ann Surg Onc. 2008.
3. Takeda et al. Measurement of cytokeratin 19 fragments as a marker of lung cancer by CYFRA21-1 enzyme immuno assay. BJC. 1995.
4. Nakata et al. Serum CYFRA21-1 (cytokeiin-19 fragments) is a useful tumour marker for detecting disease relapse and assessing treatment efficacy in breast cancer. BJC 2004.

Figure 74 (Abstract 3). Copy of abstract poster presented at the EASL monothematic conference on PSC (Copenhagen, 2009) demonstrating the diagnostic and prognostic utility of CYFRA 21-1 in patients with sporadic and PSC related BTC.

Degraded RNA isolated from bile and biliary brushings is useful for qPCR

Chapman MH[1], Andreola F[1], Cerec V[1], Sandanayake N[1], Dooley JS[2] & Pereira SP[1].

[1]The Institute of Hepatology and [2]Centre for Hepatology, Royal Free & University College London Medical School, London, UK.

Background

RNA isolated from bile or biliary brushings taken at endoscopic retrograde cholangiopancreatography (ERCP) is usually highly degraded, even when intact cells are seen on cytology. Recent advances in molecular biology now permit degraded RNA to be used in downstream applications such as real time quantitative polymerase chain reaction (qPCR) and microarray analysis.

Aims

We aimed to assess whether: (i) RNA isolated from bile or biliary brushings is useful for qPCR; and (ii) bile or x-ray contrast (iohexol [Omnipaque®]) used at the time of ERCP could be responsible for the RNA degradation observed in detached cells.

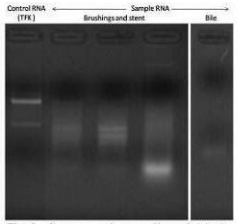


Fig 1. Agarose gel separation of intact RNA from TFK cells and degraded RNA from clinical samples of biliary brushings and bile

Methods

Fresh bile and biliary brushings were obtained at the time of ERCP or percutaneous biliary drainage. Biliary brushings were snap frozen in liquid nitrogen. Samples were stored at -80°C. For spiking experiments, TFK-1 cells (an extra-hepatic cholangiocarcinoma cell line) were incubated in 0.45μM filtered bile or Omnipaque solution (25% or 50%) for 0 to 24 hours. Total RNA was isolated using TRI reagent® (Ambion) and, to remove genomic DNA contamination, treated with DNase I (TurboDNase, Ambion) followed by spin column purification (RNeasy MinElute, Qiagen). RNA was quantified by NanoDrop ND-1000 spectrophotometer (Labtech) and its quality assessed by both gel electrophoresis and Agilent 2100 Bioanalyzer. cDNA synthesis reactions were primed with random hexamers (Superscript, invitrogen). qPCR was performed using SYBR Green dye. The expression of the housekeeping gene GAPDH (NM 002046) was assessed using four sets of primers producing amplicons of variable length (87, 131, 226 and 406bp, table 1). Amplification of CD45 [NM 080922.1] was assessed (to ascertain the degree of leukocyte RNA contamination in clinical samples) as well as expression of CK19 [NM 002276.3] and EGFR [NM 005228]) as epithelial markers.

Results

RNA isolated from bile and biliary brushings was of variable quantity (100-1500ng), poor quality (Agilent RNA Integrity Number (RIN) <5) and degraded, as assessed by both agarose gel electrophoresis and Agilent Bionalyzer (Fig. 1 and 2A, respectively). In bile and biliary brushings, RNA fragments were estimated by Agilent Bioanalyzer plots to be primarily 100 to 600nt long. Partially degraded RNA samples (*e.g.* from brushings) retained clear 28S & 18S rRNA peaks and showed presence of fragments larger in size than the ones detected in highly degraded bile samples (Fig. 2A). In contrast, RNA isolated from TFK-1 cells spiked in bile or in x-ray contrast solution (Omnipaque) for up to 6 hours was of very good quality (fig. 2).

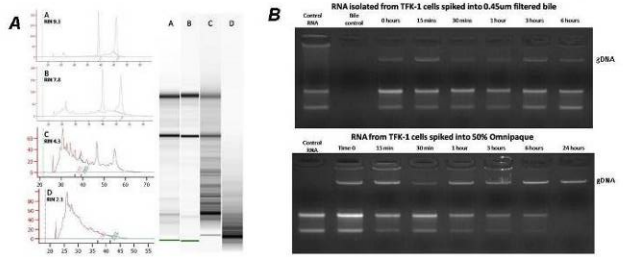
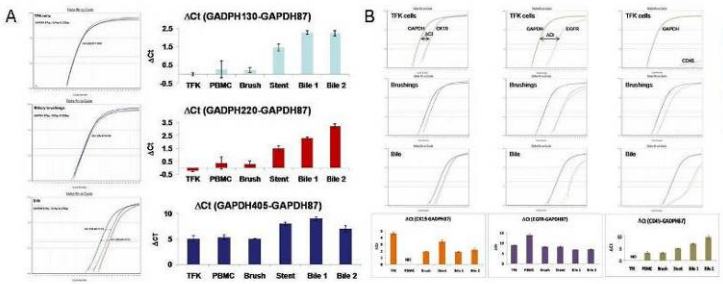


Fig 2. **Assessment of total RNA integrity.**
A) Representative Agilent bioanalyzer plots of RNA isolated from A) control TFK cells [intact], B) TFK cells incubated in bile (6 hours) [minimal degradation], C) biliary brushings, [partly degraded], and D) bile [highly degraded].
B) Gel separation of RNA isolated from TFK-1 cells spiked into filtered bile (*upper*) and Omnipaque solution (*lower*) for up to 24 hours. The genomic DNA contamination required sample clean up prior to Agilent analysis and qPCR. The RNA recovered after DNA removal was insufficient for assessment by agarose gel.

To test whether the RNA isolated from clinical samples was useful for reliably studying gene expression by qPCR, GAPDH mRNA expression was assessed using 4 different sets of primers; these were designed to amplify regions of the GAPDH transcript ranging from 87 to 406 base pairs in length. In the clinical samples the lowest Ct values (*i.e.* greatest GAPDH mRNA expression) were consistently obtained when using primers to amplify the shortest region (87bp) of the GAPDH transcript. However, the Ct values from the same samples rose significantly in a stepwise fashion when primers generating larger amplicons were used (fig 3). This effect was primarily seen in highly degraded RNA from bile with much less significant variation using partially degraded RNA for which primer amplicons of at least 226bp length appear to be satisfactory. In contrast, no shift in the Ct values was observed when assessing GAPDH mRNA expression in control TFK-1 cells, regardless of the primer set used. As described by others (Páska et al, Diag Mol Path 13[4] 2004), primers with larger amplicons (eg 406bp), result in a significant fall in amplification thresholds and are not suitable for qPCR using the SYBR Green method. When assessing other genes of interest in clinical samples, primers amplifying regions ≤100bp and the DCt method can be used to assess relative gene expression (fig 3).



Primer	Amplicon length	Forward sequence (5' → 3')	Reverse sequence (5' → 3')
GAPDH 87	87	[illegible]	[illegible]
GAPDH 131	131	[illegible]	[illegible]
GAPDH 226	226	[illegible]	[illegible]
GAPDH 406	406	[illegible]	[illegible]
CK19	79	[illegible]	[illegible]
EGFR	92	[illegible]	[illegible]
CD45	102	[illegible]	[illegible]

Table 1. qPCR primer sets

Fig 3. qPCR plots and mean ΔCt values (+/- 1 standard deviation) for A) representative samples using increasing size GAPDH primer amplicons and B) CK19, EGFR and CD45

Conclusions

Despite being highly degraded, RNA isolated from bile and biliary brushings is suitable for downstream applications such as qPCR using appropriately designed gene-specific primers amplifying regions ≤100bp. These data suggest that RNA isolated from bile and biliary brushings may also be suitable for high throughput techniques such as gene expression arrays.

This research was supported by the British Liver Trust

Figure 75 (Abstract 4). Copy of poster abstract presented at the AASLD single topic conference; Pathobiology of Biliary Epithelia and Cholangiocarcinoma, Atlanta, USA, June 2008

Whole genome RNA expression profiling of ERCP biliary brushings in biliary tract cancer.

Chapman MH[1], Andreola F[1], Sandanayake N[1], Dooley JS[2] & Pereira SP[1].

[1]The Institute of Hepatology and [2]Centre for Hepatology, Royal Free & University College London Medical School, London, UK.

Background

•Biliary tract cancer (BTC) is difficult to diagnose, in part related to the low sensitivity of brush cytology of malignant biliary strictures.

•Advances in molecular biology now permit the use of degraded RNA from formalin fixed paraffin embedded tissues for use with microarray analysis.

•We previously showed that RNA derived from biliary brushings taken at endoscopic retrograde cholangiopancreatography (ERCP, Fig 1) is degraded but can be used for gene expression profiling using quantitative real time polymerase chain reaction (qPCR)[1].

Aims

To assess the utility of microarray gene expression profiling of biliary brushings to identify novel biomarkers of biliary tract cancer (BTC).

Methods

•Samples were collected from patients with BTC (cholangiocarcinoma) and benign biliary disease (chronic pancreatitis, sphincter of Oddi dysfunction, autoimmune pancreatitis & primary sclerosing cholangitis).

•Fresh biliary brushings obtained at the time of ERCP were snap frozen in liquid nitrogen. Total RNA was isolated using TRI reagent® (Ambion), DNase I digested (TurboDNase, Ambion) followed by spin column purification (RNeasy MinElute, Qiagen).

•RNA was assessed by NanoDrop ND-1000 spectrophotometer, gel electrophoresis and Agilent 2100 Bioanalyzer.

• cDNA synthesis, amplification and labelling for microarray analysis were carried out using Nugen FFPE kits designed for fragmented RNA.

•Samples (n=10) were hybridised to Affymetrix U133 plus 2.0 GeneChips and data analysed to identify genes in BTC with a fold change of > 2.0 following correction for multiple hypothesis testing (Bonferroni Hochberg test).

•Validation of the microarray data was performed by qPCR of 43 genes of interest using RNA from the microarray set and a second fresh validation set of biliary brushings (n=12).

Results- RNA quality

•RNA was primarily of epithelial origin as shown by relative expression of CK19 and CD45 (fig 2)

•RNA isolated from biliary brushings was of variable quantity (100-1500ng), poor quality (Agilent RNA Integrity Number (RIN) <5) and degraded, with RNA fragments primarily 100 to 600nt long, as assessed by both agarose gel electrophoresis and Agilent Bionalyzer (Fig. 3).

•Quality control of the microarray data showed significant variation in RNA degradation plots, MAS5 normalisation and correlation plots necessitating exclusion of one sample (BTC) prior to data analysis.

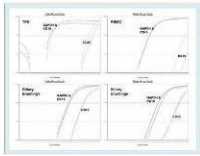


Fig 2. Representative qPCR plots for CK19 and CD45 relative to the reference gene GAPDH. The controls (top) are PBMCs for the leukocyte marker CD45 and the TFK-1 BTC cell line for the epithelial cell marker CK19. Clinical samples of biliary brushings clearly show a predominant epithelial cell source of RNA with a large shift in amplification curve for CD45 showing a minor contribution of leukocyte RNA.

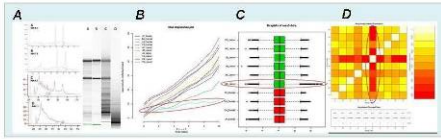


Fig 3. **Assessment of RNA quality for microarray analysis.**
A) Representative Agilent bioanalyzer plots of RNA isolated from A) control TFK cells [intact], B) control TFK cells incubated in bile (6 hours) [minimal degradation], C) biliary brushings, [partly degraded], and D) bile [highly degraded].
B) RNA degradation plots, C) MAS5 box-plots and D) MAS5 correlation plots for Affymetrix data analysis showing higher variability in one sample (circled, excluded)

Table 1. Relative gene expression of target genes normalised to 18S mRNA isolated from biliary brushings of BTC versus benign biliary disease

Gene Symbol	Gene ID	Microarray Fold change	qPCR validation (microarray set)	qPCR validation (fresh set)
MUC4	NM_018406.5_s1	8.0	19.4	10.0
MUC5AC	NM_001304359.1_s1	3.0	7.4	5.2
CD9	NM_001769.3_s1	3	2.8	1.8
MET_CHR	MET_CHR_HUMAN_s1	6	6.0	4.2
GAPDH	GAPDH_HUMAN_s1	0	1.7	0.0
ERBB2IP	ERBB2IP_NM_018308.2_s1	3	1.0	2.0
CDH17	CDH17_NM_004063.2_s1	3.0	2.4	2.4
CDH1	CDH1_NM_004360.2_s1	5.2	7.2	7.2
CLDN1	CLDN1_NM_021101.3_s1	3.0	1.1	0.7
CLDN7	CLDN7_NM_001307.3_s1	7.0	10.9	4.8
CDH3	CDH3_NM_001793.3_s1	7.0	2.6	2.6
CLDN3	CLDN3_NM_001306.3_s1	9.0	22.7	2.3
LAMB3	LAMB3_NM_000228.2_s1	9.0	10.0	3.0
MDK	MDK_NM_001012334.1_s1	16	27.6	35.0
MUC1	MUC1_NM_002456.3_s1	16	5.6	10.2
TFF3	TFF3_NM_003226.2_s1	3.7	4.6	2.6
TFF1	TFF1_NM_003225.2_s1	2.0	10.8	5.0
LAMC2	LAMC2_NM_005562.1_s1	2.5	2.8	3.0
LEF1	LEF1_NM_016269.3_s1	1.5	2.1	0.8
LIF	LIF_NM_002309.3_s1	0.7	2.5	2.0
LAP	LAP_NM_005765.2_s1	1.0	0.7	0.3
MUC2	MUC2_NM_002457.2_s1	0.7	2.5	1.4
MUC3A	MUC3A_NM_005960.3_s1	2.0	1.5	0.7
HOXA9	HOXA9_NM_152739.3_s1	2.2	1.5	2.0
TMPS	TMPS_NM_024922.2_s1	2.0	1.0	2.0
MUC13	MUC13_NM_033049.3_s1	2.0	10.4	2.0
MYC	MYC_NM_002467.3_s1	1.7	1.0	1.0
MUC7	MUC7_NM_013240.1_s1	0.4	1.0	1.0
PLAT	PLAT_NM_000930.2_s1	1.0	1.0	2.0
ERBB3	ERBB3_NM_001982.2_s1	2.0	1.0	1.0
HOXA10	HOXA10_NM_018951.2_s1	0.0	2.5	1.5
PAK1B1	PAK1B1_NM_002578.3_s1	0.0	0.0	0.0
TFF2	TFF2_NM_005423.2_s1	0	0.0	0.0
PAK1	PAK1_NM_002576.3_s1	1.0	1.0	1.0
KRAS	KRAS_NM_004985.3_s1	2.0	2.7	2.7
MUC6	MUC6_NM_005961.2_s1	0.0	1.0	0.0
SERPINB3	SERPINB3_NM_006919.2_s1	2.0	10.7	4.6
STAT1	STAT1_NM_007315.3_s1	2.0	1.4	2.0
TGFB	TGFB_NM_000660.3_s1	0.0	1.0	0.0
TM4SF4	TM4SF4_NM_004617.2_s1	1.0	0.0	0.0
TP53	TP53_NM_000546.3_s1	0.0	1.0	1.0
VEGFA	VEGFA_NM_003376.4_s1	2.0	2.0	2.0

Results- microarray analysis

•1140 upregulated and 1001 downregulated genes were identified by microarray with fold change >2 and p<0.05 (modified T test). 34,000 probe sets had non significant or fold change <2.

•The shortlist of genes further analysed by TaqMan Array qPCR was based on level of fold change, statistical significance, biological plausibility and absence of overlap in microarray gene expression plots between benign and malignant samples (fig 4).

•Validity of results was supported by identification of genes already identified as upregulated in BTC (MUC4, MUC5AC, MYC, PTK2) as well as lack of potential confounding effect of infection in BTC as evidenced by a lack of leukocyte specific markers including CD45, LSP1, CD43, CD18, CTSG, LAP and CD166.

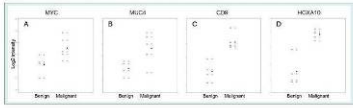


Fig 4. Microarray gene expression plots showing relative intensity of microarray signal for benign and malignant samples. Note confirmation of previously identified upregulated genes (MYC &MUC4) and lack of overlap in the potential novel biomarkers CD9 and HOXA10

Results- qPCR validation of gene expression

•TaqMan array qPCR validation of a subset of 43 upregulated genes had 79% concordance with the microarray validation set and 83% in the fresh validation set, as defined by relative gene expression of ≥2 in both sets (Figure 5 & Table 1).

•Further validation using single gene SYBR Green qPCR for MUC4, which we have previously shown to be upregulated in BTC[2], confirmed the upregulation (fold change 7.8) shown by microarray and TaqMan Array qPCR (fold change 3.2 to 29.3).

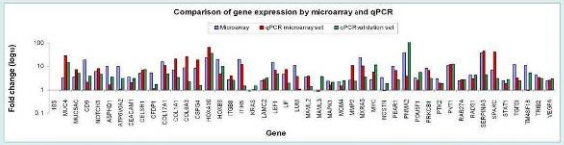


Fig 5. Comparison of relative gene expression in biliary brushings taken form patients with BTC versus benign biliary disease. Level of fold change identified by microarray and qPCR validation is shown by A) bar chart (above) and B) table (right)

Conclusions

•RNA isolated from biliary brushings is suitable for whole genome expression profiling using microarray and qPCR validation.

•Biliary brushings may be a useful resource to further investigate the pathobiology of BTC and other biliary diseases.

References 1.Chapman MH et al, *Gut* 2009;58(Suppl) 2. Mafull WR et al, *BJC* 2008;98,1675–1681

Figure 76 (Abstract 5). Copy of abstract poster presented at the AASLD meeting, Boston, 2009

9.1.3 Ethical approval, clinical databases and sample biobank

9.1.3.1 Ethical approval

Ethical approval for this work and related translational research was granted at the start of the project. The approval includes prospective sample collection of blood, bile, brushings and tissue from patients with benign and malignant hepatopancreatobiliary disorders, as well as access to similar archived samples for laboratory research. Ethical approval was granted by the Joint UCL/UCH Local Research Ethics Committee (ref 06/Q0152/106) with Site Specific Assessment at the Royal Free Hampstead NHS Trust (Dr James Dooley).

9.1.3.2 Clinical databases

Clinical databases for research purposes already in use have been maintained and new databases have also been established during the time of this project. The databases are crucial for identifying patients and data related to this and other ongoing projects. Databases include;

- Patients with primary sclerosing cholangitis
- Patients with autoimmune pancreatitis
- Patients recruited into BTC-related clinical trials e.g. PHOTOSTENT-02
- Patients with sphincter of Oddi dysfunction
- Patients discussed at the joint UCH and Royal Free Hospital HPB cancer multi-disciplinary meeting.

- Clinical samples stored in the laboratory for translational research.
- ERCP databases of archived data from the previous endoscopy software and ongoing collection with current software in use by hospital departments
- Patients listed as having had biliary brush cytology at UCH.

9.1.3.3 Biobank of clinical samples

A standard operating procedure is in place to optimise and standardise clinical sample collection. Following informed patient consent, patient samples are collected, processed, separated into multiple cryotubes and stored at -80^{o}C in freezers in our laboratory at the Institute of Hepatology. Samples are stored racked and dated in colour coded boxes for convenience. Samples are coded for anonymity and a database of stored samples is used to identify samples and record relevant basic clinical information.

9.1.4 Real Time quantitative PCR – calculation of relative gene expression using the comparative cycle threshold (ΔCt) method

Calculation of relative gene expression using qPCR data is performed using the comparative Ct method first described for use in qPCR by Livak and Schmittgen (Livak and Schmittgen 2001)

Calculation of relative gene expression using qPCR requires a stably expressed 'house keeping' gene (e.g. GAPDH, ACTB, 18SRNA) for reference and normalization. All cells from all tissues should ideally have the same expression of the reference gene in health and disease. The relative (or comparative) expression (ΔCt) of a gene of interest is compared to the reference gene using the equation:

$\Delta Ct = Ct_{\text{target gene}} - Ct_{\text{reference gene}}$

The relative gene expression of one sample (e.g. cancer) compared to another (e.g. benign) is calculated and expressed using the ΔΔCt method:

$\Delta\Delta Ct = \Delta Ct_{\text{sample}} - \Delta Ct_{\text{control}}$

Relative gene expression = $2^{-\Delta\Delta Ct}$

9.1.5 Cholangiocarcinoma Staging

The staging system most commonly used in BTC is the TNM staging outlined as follows;

"The American Joint Committee on Cancer guidelines in the *AJCC Cancer Staging Manual, Fifth Edition,* following the tumour, node, and metastasis (TNM) classification system, with depth of tumour penetration and regional spread defined pathologically, should be followed."

T - Primary tumour
TX - Primary tumour cannot be assessed
T0 - No evidence of primary tumour
TIS - Carcinoma in situ
T1a - Tumour invades mucosa
T1b - Tumour invades muscularis
T2 - Tumour invades perimuscular connective tissue
T3 - Tumour invades liver, gallbladder, duodenum, stomach, pancreas, or colon

N - Regional lymph nodes
NX - Regional lymph nodes cannot be assessed
N0 - No metastases in regional lymph nodes
N1 - Metastases in cystic duct or pericholedochal or hilar lymph nodes of hepatoduodenal ligament
N2 - Metastases in peripancreatic (head only), periduodenal, posterior pancreatoduodenal, periportal, celiac, or superior mesenteric regional lymph nodes

M - Metastasis
MX - Presence of metastases cannot be assessed
M0 - No distant metastases
M1 - Distant metastases (includes lymph node metastases beyond N2)

TNM groupings by stage

Stage 0 - TIS N0 M0
Stage I - T1 N0 M0
Stage II - T2 N0 M0
Stage III - T1-2 N1-2 M0
Stage IVa - T3 N0-2 M0
Stage IVb - T1-3 N0-2 M1

9.1.6 Techniques used in this thesis and details of assistance given

All laboratory techniques performed by myself were done following instruction and supervision of Dr Fausto Andreola, Post-Doctoral Scientist. His help and teaching is particularly appreciated.

- Conception of ideas and design of experiments

Myself with guidance of Dr Stephen Pereira, Dr Fausto Andreola and Dr James Dooley

- Ethical and R&D approval

Myself with guidance of Dr Stephen Pereira and Dr James Dooley (Royal Free Hospital)

- Sample collection, storage and maintenance of databases

Myself, Dr Neomal Sandanayake, Robert Tidswell.

Biliary brushings were taken by other clinicians at the time of ERCP procedures (Dr Stephen Pereira, Dr George Webster, Dr Adrian Hatfield & Dr James Dooley (Royal Free Hospital))

- RNA isolation and assessment of quality

Myself with guidance of Dr Fausto Andreola

- cDNA synthesis

Myself with guidance of Dr Fausto Andreola

- Standard PCR

Myself with guidance of Dr Fausto Andreola

- Single gene qPCR (SYBR Green and TaqMan)

Myself with guidance of Dr Fausto Andreola. Experiments with CD9 were done also with the help of Robert Tidswell (iBSc student)

- Mutliple gene qPCR using the TaqMan Array platform

These were performed at the CRUK Lincolns Inn Fields laboratories by Alvin Lee and myself with guidance of Dr Charles Swanton (CRUK)

- qPCR data analysis

Myself and Robert Tidswell with guidance of Dr Fausto Andreola

- Microarray experiments

Performed as a paid service by Scientific Support Services, UCL

- Microarray data analysis

Done as a paid service by Dr Sonia Shah, Bloomsbury Centre for Bioinformatics, UCL

- Tissue culture

Cells from tissue culture were provided by Dr Virginie Cerec

- Gel electrophoresis

Myself with guidance of Dr Fausto Andreola

- Western blot

Myself with guidance of Dr Virginie Cerec and Dr Fausto Andreola

- ELISA

Myself with guidance of Dr Fausto Andreola

- Immunohistochemistry

Done as a paid service by UCL Advanced Diagnostics (Philipa Munson).

- Reporting of immunohistochemistry slides

Dr Maesha Deheragoda, Department of Pathology, UCL.

- Laser Capture Microdissection

Myself with instruction and supervison by Dr Simon Leedham, CRUK

- Statistical analysis

Myself with guidance of Dr Fausto Andreola and Dr Dipok Dhar

- Preparation of images, tables and figures

Myself

10 References

Ahmetoglu, A., P. Kosucu, S. Kul, H. Dinc, A. Sari, M. Arslan, E. Alhan and H. R. Gumele (2004). "MDCT cholangiography with volume rendering for the assessment of patients with biliary obstruction." AJR Am J Roentgenol **183**(5): 1327-1332.

Ahrendt, S. A., H. A. Pitt, A. Nakeeb, A. S. Klein, K. D. Lillemoe, A. N. Kalloo and J. L. Cameron (1999). "Diagnosis and management of cholangiocarcinoma in primary sclerosing cholangitis." J Gastrointest Surg **3**(4): 357-367; discussion 367-358.

Allard, W. J., J. Matera, M. C. Miller, M. Repollet, M. C. Connelly, C. Rao, A. G. Tibbe, J. W. Uhr and L. W. Terstappen (2004). "Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases." Clin Cancer Res **10**(20): 6897-6904.

Allinen, M., R. Beroukhim, L. Cai, C. Brennan, J. Lahti-Domenici, H. Huang, D. Porter, M. Hu, L. Chin, A. Richardson, S. Schnitt, W. R. Sellers and K. Polyak (2004). "Molecular characterization of the tumor microenvironment in breast cancer." Cancer Cell **6**(1): 17-32.

Aloia, T. A., C. Charnsangavej, S. Faria, D. Ribero, E. K. Abdalla, J. N. Vauthey and S. A. Curley (2007). "High-resolution computed tomography accurately predicts resectability in hilar cholangiocarcinoma." Am J Surg **193**(6): 702-706.

Anderson, C. D., M. H. Rice, C. W. Pinson, W. C. Chapman, R. S. Chari and D. Delbeke (2004). "Fluorodeoxyglucose PET imaging in the evaluation

of gallbladder carcinoma and cholangiocarcinoma." J Gastrointest Surg **8**(1): 90-97.

Andreadis, C., S. Touloupidis, G. Galaktidou, A. H. Kortsaris, A. Boutis and D. Mouratidou (2005). "Serum CYFRA 21-1 in patients with invasive bladder cancer and its relevance as a tumor marker during chemotherapy." J Urol **174**(5): 1771-1775; discussion 1775-1776.

Angulo, P., D. H. Pearce, C. D. Johnson, J. J. Henry, N. F. LaRusso, B. T. Petersen and K. D. Lindor (2000). "Magnetic resonance cholangiography in patients with biliary disease: its role in primary sclerosing cholangitis." J Hepatol **33**(4): 520-527.

Antonov, J., D. R. Goldstein, A. Oberli, A. Baltzer, M. Pirotta, A. Fleischmann, H. J. Altermatt and R. Jaggi (2005). "Reliable gene expression measurements from degraded RNA by quantitative real-time PCR depend on short amplicons and a proper normalization." Lab Invest **85**(8): 1040-1050.

Ayaru, L., K. Stoeber, G. J. Webster, A. R. Hatfield, A. Wollenschlaeger, O. Okoturo, M. Rashid, G. Williams and S. P. Pereira (2008). "Diagnosis of pancreaticobiliary malignancy by detection of minichromosome maintenance protein 5 in bile aspirates." Br J Cancer **98**(9): 1548-1554.

Baba, N., H. Kobashi, K. Yamamoto, R. Terada, T. Suzuki, T. Hakoda, N. Okano, N. Shimada, S. Fujioka, Y. Iwasaki and Y. Shiratori (2006). "Gene expression profiling in biliary epithelial cells of primary biliary cirrhosis using laser capture microdissection and cDNA microarray." Transl Res **148**(3): 103-113.

Banyard, J., L. Bao, M. D. Hofer, D. Zurakowski, K. A. Spivey, A. S. Feldman, L. M. Hutchinson, R. Kuefer, M. A. Rubin and B. R. Zetter (2007). "Collagen XXIII expression is associated with prostate cancer recurrence and distant metastases." Clin Cancer Res **13**(9): 2634-2642.

Baron, T. H., G. C. Harewood, A. Rumalla, N. L. Pochron, L. M. Stadheim, G. J. Gores, T. M. Therneau, P. C. De Groen, T. J. Sebo, D. R. Salomao and B. R. Kipp (2004). "A prospective comparison of digital image analysis and routine cytology for the identification of malignancy in biliary tract strictures." Clin Gastroenterol Hepatol **2**(3): 214-219.

Beck-Engeser, G. B., A. M. Lum, K. Huppi, N. J. Caplen, B. B. Wang and M. Wabl (2008). "Pvt1-encoded microRNAs in oncogenesis." Retrovirology **5**: 4.

Becker, N. S., J. A. Rodriguez, N. R. Barshes, C. A. O'Mahony, J. A. Goss and T. A. Aloia (2008). "Outcomes analysis for 280 patients with cholangiocarcinoma treated with liver transplantation over an 18-year period." J Gastrointest Surg **12**(1): 117-122.

Benedetti, A., D. Alvaro, C. Bassotti, A. Gigliozzi, G. Ferretti, T. La Rosa, A. Di Sario, L. Baiocchi and A. M. Jezequel (1997). "Cytotoxicity of bile salts against biliary epithelium: a study in isolated bile ductule fragments and isolated perfused rat liver." Hepatology **26**(1): 9-21.

Bergquist, A., A. Ekbom, R. Olsson, D. Kornfeldt, L. Loof, A. Danielsson, R. Hultcrantz, S. Lindgren, H. Prytz, H. Sandberg-Gertzen, S. Almer, F. Granath and U. Broome (2002). "Hepatic and extrahepatic

malignancies in primary sclerosing cholangitis." J Hepatol **36**(3): 321-327.

Berstad, A. E., L. Aabakken, H. J. Smith, S. Aasen, K. M. Boberg and E. Schrumpf (2006). "Diagnostic accuracy of magnetic resonance and endoscopic retrograde cholangiography in primary sclerosing cholangitis." Clin Gastroenterol Hepatol **4**(4): 514-520.

Bidard, F. C., C. Mathiot, S. Delaloge, E. Brain, S. Giachetti, P. de Cremoux, M. Marty and J. Y. Pierga (2009). "Single circulating tumor cell detection and overall survival in nonmetastatic breast cancer." Ann Oncol.

Bjornsson, E., A. Kilander and R. Olsson (1999). "CA 19-9 and CEA are unreliable markers for cholangiocarcinoma in patients with primary sclerosing cholangitis." Liver **19**(6): 501-508.

Bjornsson, E., R. Olsson, A. Bergquist, S. Lindgren, B. Braden, R. W. Chapman, K. M. Boberg and P. Angulo (2008). "The natural history of small-duct primary sclerosing cholangitis." Gastroenterology **134**(4): 975-980.

Blechacz, B. and G. J. Gores (2008). "Cholangiocarcinoma: advances in pathogenesis, diagnosis, and treatment." Hepatology **48**(1): 308-321.

Blechacz, B. R., W. Sanchez and G. J. Gores (2009). "A conceptual proposal for staging ductal cholangiocarcinoma." Curr Opin Gastroenterol **25**(3): 238-239.

Boberg, K. M., A. Bergquist, S. Mitchell, A. Pares, F. Rosina, U. Broome, R. Chapman, O. Fausa, T. Egeland, G. Rocca and E. Schrumpf (2002).

"Cholangiocarcinoma in primary sclerosing cholangitis: risk factors and clinical presentation." Scand J Gastroenterol **37**(10): 1205-1211.

Boberg, K. M., P. Jebsen, O. P. Clausen, A. Foss, L. Aabakken and E. Schrumpf (2006). "Diagnostic benefit of biliary brush cytology in cholangiocarcinoma in primary sclerosing cholangitis." J Hepatol **45**(4): 568-574.

Boheler, K. R., M. Volkova, C. Morrell, R. Garg, Y. Zhu, K. Margulies, A. M. Seymour and E. G. Lakatta (2003). "Sex- and age-dependent human transcriptome variability: implications for chronic heart failure." Proc Natl Acad Sci U S A **100**(5): 2754-2759.

Brechot, J. M., S. Chevret, J. Nataf, C. Le Gall, J. Fretault, J. Rochemaure and C. Chastang (1997). "Diagnostic and prognostic value of Cyfra 21-1 compared with other tumour markers in patients with non-small cell lung cancer: a prospective study of 116 patients." Eur J Cancer **33**(3): 385-391.

Buffet, C., C. Fourre, C. Altman, F. Prat, J. Fritsch, A. Choury, M. J. Briantais, A. Desgrez and J. P. Etienne (1996). "Bile levels of carcino-embryonic antigen in patients with hepatopancreatobiliary disease." Eur J Gastroenterol Hepatol **8**(2): 131-134.

Burak, K., P. Angulo, T. M. Pasha, K. Egan, J. Petz and K. D. Lindor (2004). "Incidence and risk factors for cholangiocarcinoma in primary sclerosing cholangitis." Am J Gastroenterol **99**(3): 523-526.

Canales, R. D., Y. Luo, J. C. Willey, B. Austermiller, C. C. Barbacioru, C. Boysen, K. Hunkapiller, R. V. Jensen, C. R. Knight, K. Y. Lee, Y. Ma, B. Maqsodi, A. Papallo, E. H. Peters, K. Poulter, P. L. Ruppel, R. R.

Samaha, L. Shi, W. Yang, L. Zhang and F. M. Goodsaid (2006). "Evaluation of DNA microarray results with quantitative gene expression platforms." Nat Biotechnol **24**(9): 1115-1122.

Cha, J. H., J. K. Han, T. K. Kim, A. Y. Kim, S. J. Park, B. I. Choi, K. S. Suh, S. W. Kim and M. C. Han (2000). "Preoperative evaluation of Klatskin tumor: accuracy of spiral CT in determining vascular invasion as a sign of unresectability." Abdom Imaging **25**(5): 500-507.

Cha, J. M., M. H. Kim, S. K. Lee, D. W. Seo, S. S. Lee, J. H. Lee, S. G. Lee and S. J. Jang (2006). "Clinicopathological review of 61 patients with early bile duct cancer." Clin Oncol (R Coll Radiol) **18**(9): 669-677.

Chalasani, N., A. Baluyut, A. Ismail, A. Zaman, G. Sood, R. Ghalib, T. M. McCashland, K. R. Reddy, X. Zervos, M. A. Anbari and H. Hoen (2000). "Cholangiocarcinoma in patients with primary sclerosing cholangitis: a multicenter case-control study." Hepatology **31**(1): 7-11.

Chapman, M. H. and S. P. Pereira (2009). "Cholangiocarcinoma: improving biliary drainage with PDT." Photodiagnosis Photodyn Ther **6**(2): 82-83.

Chapman, M. H., J. Witmann, G. J. Webster, A. H. Hatfield and S. P. Pereira (2008). "A single centre experience of dominant strictures and cholangiocarcinoma in orimary sclerosing cholangitis " Bile Acid Biology and Therapeutics, Amsterdam **Abstract**.

Chapman, R. W., B. A. Arborgh, J. M. Rhodes, J. A. Summerfield, R. Dick, P. J. Scheuer and S. Sherlock (1980). "Primary sclerosing cholangitis: a review of its clinical features, cholangiography, and hepatic histology." Gut **21**(10): 870-877.

Charatcharoenwitthaya, P., F. B. Enders, K. C. Halling and K. D. Lindor (2008). "Utility of serum tumor markers, imaging, and biliary cytology for detecting cholangiocarcinoma in primary sclerosing cholangitis." Hepatology **48**(4): 1106-1117.

Chen, C. Y., S. C. Shiesh, H. C. Tsao and X. Z. Lin (2002). "The assessment of biliary CA 125, CA 19-9 and CEA in diagnosing cholangiocarcinoma--the influence of sampling time and hepatolithiasis." Hepatogastroenterology **49**(45): 616-620.

Chen, Y. K. and D. K. Pleskow (2007). "SpyGlass single-operator peroral cholangiopancreatoscopy system for the diagnosis and therapy of bile-duct disorders: a clinical feasibility study (with video)." Gastrointest Endosc **65**(6): 832-841.

Cheng, W., J. Liu, H. Yoshida, D. Rosen and H. Naora (2005). "Lineage infidelity of epithelial ovarian cancers is controlled by HOX genes that specify regional identity in the reproductive tract." Nat Med **11**(5): 531-537.

Chuaysri, C., P. Thuwajit, A. Paupairoj, S. Chau-In, T. Suthiphongchai and C. Thuwajit (2009). "Alpha-smooth muscle actin-positive fibroblasts promote biliary cell proliferation and correlate with poor survival in cholangiocarcinoma." Oncol Rep **21**(4): 957-969.

Chung, J. Y., T. Braunschweig and S. M. Hewitt (2006). "Optimization of recovery of RNA from formalin-fixed, paraffin-embedded tissue." Diagn Mol Pathol **15**(4): 229-236.

Connor, S., E. Barron, D. N. Redhead, H. Ireland, K. K. Madhavan, R. W. Parks and O. J. Garden (2007). "Palliation for suspected unresectable hilar cholangiocarcinoma." Eur J Surg Oncol **33**(3): 341-345.

Coudry, R. A., S. I. Meireles, R. Stoyanova, H. S. Cooper, A. Carpino, X. Wang, P. F. Engstrom and M. L. Clapper (2007). "Successful application of microarray technology to microdissected formalin-fixed, paraffin-embedded tissue." J Mol Diagn **9**(1): 70-79.

Cristofanilli, M., G. T. Budd, M. J. Ellis, A. Stopeck, J. Matera, M. C. Miller, J. M. Reuben, G. V. Doyle, W. J. Allard, L. W. Terstappen and D. F. Hayes (2004). "Circulating tumor cells, disease progression, and survival in metastatic breast cancer." N Engl J Med **351**(8): 781-791.

Cronin, M., M. Pho, D. Dutta, J. C. Stephans, S. Shak, M. C. Kiefer, J. M. Esteban and J. B. Baker (2004). "Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay." Am J Pathol **164**(1): 35-42.

Davidson, B., N. Varsamidakis, J. Dooley, A. Deery, R. Dick, T. Kurzawinski and K. Hobbs (1992). "Value of exfoliative cytology for investigating bile duct strictures." Gut **33**(10): 1408-1411.

de Bellis, M., S. Sherman, E. L. Fogel, H. Cramer, J. Chappo, L. McHenry, Jr., J. L. Watkins and G. A. Lehman (2002). "Tissue sampling at ERCP in suspected malignant biliary strictures (Part 2)." Gastrointest Endosc **56**(5): 720-730.

de Groen, P. C., G. J. Gores, N. F. LaRusso, L. L. Gunderson and D. M. Nagorney (1999). "Biliary tract cancers." N Engl J Med **341**(18): 1368-1378.

Desa, L. A., A. B. Akosa, S. Lazzara, P. Domizio, T. Krausz and I. S. Benjamin (1991). "Cytodiagnosis in the management of extrahepatic biliary stricture." Gut **32**(10): 1188-1191.

Dhanasekaran, S. M., T. R. Barrette, D. Ghosh, R. Shah, S. Varambally, K. Kurachi, K. J. Pienta, M. A. Rubin and A. M. Chinnaiyan (2001). "Delineation of prognostic biomarkers in prostate cancer." Nature **412**(6849): 822-826.

Domagk, D., J. Wessling, P. Reimer, L. Hertel, C. Poremba, N. Senninger, A. Heinecke, W. Domschke and J. Menzel (2004). "Endoscopic retrograde cholangiopancreatography, intraductal ultrasonography, and magnetic resonance cholangiopancreatography in bile duct strictures: a prospective comparison of imaging diagnostics with histopathological correlation." Am J Gastroenterol **99**(9): 1684-1689.

Dressman, H. K., C. Hans, A. Bild, J. A. Olson, E. Rosen, P. K. Marcom, V. B. Liotcheva, E. L. Jones, Z. Vujaskovic, J. Marks, M. W. Dewhirst, M. West, J. R. Nevins and K. Blackwell (2006). "Gene expression profiles of multiple breast cancer phenotypes and response to neoadjuvant chemotherapy." Clin Cancer Res **12**(3 Pt 1): 819-826.

Duffy, M. J. (2001). "Carcinoembryonic antigen as a marker for colorectal cancer: is it clinically useful?" Clin Chem **47**(4): 624-630.

Eloubeidi, M. A., V. K. Chen, N. C. Jhala, I. E. Eltoum, D. Jhala, D. C. Chhieng, S. A. Syed, S. M. Vickers and C. Mel Wilcox (2004).

"Endoscopic ultrasound-guided fine needle aspiration biopsy of suspected cholangiocarcinoma." Clin Gastroenterol Hepatol **2**(3): 209-213.

Enjoji, M., M. Nakamuta and H. Nawata (2004). "Oncogenes in tumors of gallbladder and biliary tract." Fukuoka Igaku Zasshi **95**(2): 31-35.

Erdogan, D., J. J. Kloek, F. J. ten Kate, E. A. Rauws, O. R. Busch, D. J. Gouma and T. M. van Gulik (2008). "Immunoglobulin G4-related sclerosing cholangitis in patients resected for presumed malignant bile duct strictures." Br J Surg **95**(6): 727-734.

Ernst, O., T. Asselah, G. Sergent, M. Calvo, N. Talbodec, J. C. Paris and C. L'Hermine (1998). "MR cholangiography in primary sclerosing cholangitis." AJR Am J Roentgenol **171**(4): 1027-1030.

Farley, D. R., A. L. Weaver and D. M. Nagorney (1995). ""Natural history" of unresected cholangiocarcinoma: patient outcome after noncurative intervention." Mayo Clin Proc **70**(5): 425-429.

Farrell, R. J., A. K. Jain, S. L. Brandwein, H. Wang, R. Chuttani and D. K. Pleskow (2001). "The combination of stricture dilation, endoscopic needle aspiration, and biliary brushings significantly improves diagnostic yield from malignant bile duct strictures." Gastrointest Endosc **54**(5): 587-594.

Fava, G., M. Marzioni, A. Benedetti, S. Glaser, S. DeMorrow, H. Francis and G. Alpini (2007). "Molecular pathology of biliary tract cancers." Cancer Lett **250**(2): 155-167.

Feldmann, G., J. Nattermann, H. D. Nischalke, M. Gorschluter, T. Kuntzen, G. Ahlenstiel, T. Gerhardt, M. Wolff, T. Sauerbruch, U. Spengler and F. L.

Dumoulin (2006). "Detection of human aspartyl (asparaginyl) beta-hydroxylase and homeobox B7 mRNA in brush cytology specimens from patients with bile duct cancer." Endoscopy **38**(6): 604-609.

Fevery, J., O. Buchel, F. Nevens, C. Verslype, S. Stroobants and W. Van Steenbergen (2005). "Positron emission tomography is not a reliable method for the early diagnosis of cholangiocarcinoma in patients with primary sclerosing cholangitis." J Hepatol **43**(2): 358-360.

Fisher, A., N. D. Theise, A. Min, E. Mor, S. Emre, A. Pearl, M. E. Schwartz, C. M. Miller and P. A. Sheiner (1995). "CA19-9 does not predict cholangiocarcinoma in patients with primary sclerosing cholangitis undergoing liver transplantation." Liver Transpl Surg **1**(2): 94-98.

Fizazi, K., L. Morat, L. Chauveinc, D. Prapotnich, R. De Crevoisier, B. Escudier, X. Cathelineau, F. Rozet, G. Vallancien, L. Sabatier and J. C. Soria (2007). "High detection rate of circulating tumor cells in blood of patients with prostate cancer using telomerase activity." Ann Oncol **18**(3): 518-521.

Flamm, C. R., D. H. Mark and N. Aronson (2002). "Evidence-based assessment of ERCP approaches to managing pancreaticobiliary malignancies." Gastrointest Endosc **56**(6 Suppl): S218-225.

Fleming, K. A., K. M. Boberg, H. Glaumann, A. Bergquist, D. Smith and O. P. Clausen (2001). "Biliary dysplasia as a marker of cholangiocarcinoma in primary sclerosing cholangitis." J Hepatol **34**(3): 360-365.

Foutch, P. G., D. M. Kerr, J. R. Harlan and T. D. Kummet (1991). "A prospective, controlled analysis of endoscopic cytotechniques for

diagnosis of malignant biliary strictures." Am J Gastroenterol **86**(5): 577-580.

Fritscher-Ravens, A., D. C. Broering, W. T. Knoefel, X. Rogiers, P. Swain, F. Thonke, C. Bobrowski, T. Topalidis and N. Soehendra (2004). "EUS-guided fine-needle aspiration of suspected hilar cholangiocarcinoma in potentially operable patients with negative brush cytology." Am J Gastroenterol **99**(1): 45-51.

Fukuda, Y., T. Tsuyuguchi, Y. Sakai, S. Tsuchiya and H. Saisyo (2005). "Diagnostic utility of peroral cholangioscopy for various bile-duct lesions." Gastrointest Endosc **62**(3): 374-382.

Gandemer, V., A. G. Rio, M. de Tayrac, V. Sibut, S. Mottier, B. Ly Sunnaram, C. Henry, A. Monnier, C. Berthou, E. Le Gall, A. Le Treut, C. Schmitt, J. Y. Le Gall, J. Mosser and M. D. Galibert (2007). "Five distinct biological processes and 14 differentially expressed genes characterize TEL/AML1-positive leukemia." BMC Genomics **8**: 385.

Garber, M. E., O. G. Troyanskaya, K. Schluens, S. Petersen, Z. Thaesler, M. Pacyna-Gengelbach, M. van de Rijn, G. D. Rosen, C. M. Perou, R. I. Whyte, R. B. Altman, P. O. Brown, D. Botstein and I. Petersen (2001). "Diversity of gene expression in adenocarcinoma of the lung." Proc Natl Acad Sci U S A **98**(24): 13784-13789.

Gervasoni, A., R. M. Monasterio Munoz, G. S. Wengler, A. Rizzi, A. Zaniboni and O. Parolini (2008). "Molecular signature detection of circulating tumor cells using a panel of selected genes." Cancer Lett **263**(2): 267-279.

Glasbrenner, B., M. Ardan, W. Boeck, G. Preclik, P. Moller and G. Adler (1999). "Prospective evaluation of brush cytology of biliary strictures during endoscopic retrograde cholangiopancreatography." Endoscopy **31**(9): 712-717.

Grubman, S. A., R. D. Perrone, D. W. Lee, S. L. Murray, L. C. Rogers, L. I. Wolkoff, A. E. Mulberg, V. Cherington and D. M. Jefferson (1994). "Regulation of intracellular pH by immortalized human intrahepatic biliary epithelial cell lines." Am J Physiol **266**(6 Pt 1): G1060-1070.

Gruenberger, B., J. Schueller, U. Heubrandtner, F. Wrba, D. Tamandl, K. Kaczirek, R. Roka, S. Freimann-Pircher and T. Gruenberger (2010). "Cetuximab, gemcitabine, and oxaliplatin in patients with unresectable advanced or metastatic biliary tract cancer: a phase 2 study." Lancet Oncol **11**(12): 1142-1148.

Guan, Y., W. L. Kuo, J. L. Stilwell, H. Takano, A. V. Lapuk, J. Fridlyand, J. H. Mao, M. Yu, M. A. Miller, J. L. Santos, S. E. Kalloger, J. W. Carlson, D. G. Ginzinger, S. E. Celniker, G. B. Mills, D. G. Huntsman and J. W. Gray (2007). "Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer." Clin Cancer Res **13**(19): 5745-5755.

Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.

Hansel, D. E., A. Rahman, M. Hidalgo, P. J. Thuluvath, K. D. Lillemoe, R. Shulick, J. L. Ku, J. G. Park, K. Miyazaki, R. Ashfaq, Wistuba, II, R. Varma, L. Hawthorne, J. Geradts, P. Argani and A. Maitra (2003). "Identification of novel cellular targets in biliary tract cancers using global gene expression technology." Am J Pathol **163**(1): 217-229.

Harewood, G. C., T. H. Baron, L. M. Stadheim, B. R. Kipp, T. J. Sebo and D. R. Salomao (2004). "Prospective, blinded assessment of factors influencing the accuracy of biliary cytology interpretation." Am J Gastroenterol **99**(8): 1464-1469.

Hass, H. G., O. Nehls, J. Jobst, A. Frilling, U. Vogel and S. Kaiser (2008). "Identification of osteopontin as the most consistently over-expressed gene in intrahepatic cholangiocarcinoma: Detection by oligonucleotide microarray and real-time PCR analysis." World J Gastroenterol **14**(16): 2501-2510.

Hoffmann, K., C. Kerner, W. Wilfert, M. Mueller, J. Thiery, J. Hauss and H. Witzigmann (2007). "Detection of disseminated pancreatic cells by amplification of cytokeratin-19 with quantitative RT-PCR in blood, bone marrow and peritoneal lavage of pancreatic carcinoma patients." World J Gastroenterol **13**(2): 257-263.

Hoshida, Y., A. Villanueva, M. Kobayashi, J. Peix, D. Y. Chiang, A. Camargo, S. Gupta, J. Moore, M. J. Wrobel, J. Lerner, M. Reich, J. A. Chan, J. N. Glickman, K. Ikeda, M. Hashimoto, G. Watanabe, M. G. Daidone, S. Roayaie, M. Schwartz, S. Thung, H. B. Salvesen, S. Gabriel, V. Mazzaferro, J. Bruix, S. L. Friedman, H. Kumada, J. M. Llovet and T. R. Golub (2008). "Gene expression in fixed tissues and outcome in hepatocellular carcinoma." N Engl J Med **359**(19): 1995-2004.

Hultcrantz, R., R. Olsson, A. Danielsson, G. Jarnerot, L. Loof, B. O. Ryden, B. Wahren and U. Broome (1999). "A 3-year prospective study on serum tumor markers used for detecting cholangiocarcinoma in patients with primary sclerosing cholangitis." J Hepatol **30**(4): 669-673.

Huntsman, D. G., S. F. Chin, M. Muleris, S. J. Batley, V. P. Collins, L. M. Wiedemann, S. Aparicio and C. Caldas (1999). "MLL2, the second human homolog of the Drosophila trithorax gene, maps to 19q13.1 and is amplified in solid tumor cell lines." Oncogene **18**(56): 7975-7984.

Hwang, R. F., T. Moore, T. Arumugam, V. Ramachandran, K. D. Amos, A. Rivera, B. Ji, D. B. Evans and C. D. Logsdon (2008). "Cancer-associated stromal fibroblasts promote pancreatic tumor progression." Cancer Res **68**(3): 918-926.

Iacobuzio-Donahue, C. A., A. Maitra, G. L. Shen-Ong, T. van Heek, R. Ashfaq, R. Meyer, K. Walter, K. Berg, M. A. Hollingsworth, J. L. Cameron, C. J. Yeo, S. E. Kern, M. Goggins and R. H. Hruban (2002). "Discovery of novel tumor markers of pancreatic cancer using global gene expression technology." Am J Pathol **160**(4): 1239-1249.

Isomoto, H., S. Kobayashi, N. W. Werneburg, S. F. Bronk, M. E. Guicciardi, D. A. Frank and G. J. Gores (2005). "Interleukin 6 upregulates myeloid cell leukemia-1 expression through a STAT3 pathway in cholangiocarcinoma cells." Hepatology **42**(6): 1329-1338.

Ito, F., R. Agni, R. J. Rettammel, M. J. Been, C. S. Cho, D. M. Mahvi, L. F. Rikkers and S. M. Weber (2008). "Resection of hilar cholangiocarcinoma: concomitant liver resection decreases hepatic recurrence." Ann Surg **248**(2): 273-279.

Jailwala, J., E. L. Fogel, S. Sherman, K. Gottlieb, J. Flueckiger, L. G. Bucksot and G. A. Lehman (2000). "Triple-tissue sampling at ERCP in malignant biliary obstruction." Gastrointest Endosc **51**(4 Pt 1): 383-390.

Jarnagin, W. R., Y. Fong, R. P. DeMatteo, M. Gonen, E. C. Burke, B. J. Bodniewicz, B. M. Youssef, D. Klimstra and L. H. Blumgart (2001). "Staging, resectability, and outcome in 225 patients with hilar cholangiocarcinoma." Ann Surg **234**(4): 507-517; discussion 517-509.

Jinawath, N., Y. Chamgramol, Y. Furukawa, K. Obama, T. Tsunoda, B. Sripa, C. Pairojkul and Y. Nakamura (2006). "Comparison of gene expression profiles between Opisthorchis viverrini and non-Opisthorchis viverrini associated human intrahepatic cholangiocarcinoma." Hepatology **44**(4): 1025-1038.

John, A. R., K. S. Haghighi, P. Taniere, M. E. Esmat, Y. M. Tan and S. R. Bramhall (2006). "Is a raised CA 19-9 level diagnostic for a cholangiocarcinoma in patients with no history of sclerosing cholangitis ?" Dig Surg **23**(5-6): 319-324.

Jones, S., X. Zhang, D. W. Parsons, J. C. Lin, R. J. Leary, P. Angenendt, P. Mankoo, H. Carter, H. Kamiyama, A. Jimeno, S. M. Hong, B. Fu, M. T. Lin, E. S. Calhoun, M. Kamiyama, K. Walter, T. Nikolskaya, Y. Nikolsky, J. Hartigan, D. R. Smith, M. Hidalgo, S. D. Leach, A. P. Klein, E. M. Jaffee, M. Goggins, A. Maitra, C. Iacobuzio-Donahue, J. R. Eshleman, S. E. Kern, R. H. Hruban, R. Karchin, N. Papadopoulos, G. Parmigiani, B. Vogelstein, V. E. Velculescu and K. W. Kinzler (2008). "Core signaling pathways in human pancreatic cancers revealed by global genomic analyses." Science **321**(5897): 1801-1806.

Ju, Y. M., M. H. Kim, S. K. Lee, D. W. Seo, Y. I. Min and J. Y. Kim (2002). "Comparative cytotoxicity of low-osmolar nonionic and high-osmolar

ionic contrast media to dog gallbladder epithelial cells." Gastrointest Endosc **55**(3): 382-386.

Kashihara, T., A. Ohki, T. Kobayashi, T. Sato, H. Nishizawa, K. Ogawa, H. Tako, F. Kawakami, M. Tsuji and K. Tamaoka (1998). "Intrahepatic cholangiocarcinoma with increased serum CYFRA 21-1 level." J Gastroenterol **33**(3): 447-453.

Ker, C. G., J. S. Chen, K. T. Lee, P. C. Sheen and C. C. Wu (1991). "Assessment of serum and bile levels of CA19-9 and CA125 in cholangitis and bile duct carcinoma." J Gastroenterol Hepatol **6**(5): 505-508.

Khan, S. A., B. R. Davidson, R. Goldin, S. P. Pereira, W. M. Rosenberg, S. D. Taylor-Robinson, A. V. Thillainayagam, H. C. Thomas, M. R. Thursz and H. Wasan (2002). "Guidelines for the diagnosis and treatment of cholangiocarcinoma: consensus document." Gut **51 Suppl 6**: VI1-9.

Khan, S. A., S. D. Taylor-Robinson, M. B. Toledano, A. Beck, P. Elliott and H. C. Thomas (2002). "Changing international trends in mortality rates for liver, biliary and pancreatic tumours." J Hepatol **37**(6): 806-813.

Kiguchi, K., L. Ruffino, T. Kawamoto, T. Ajiki and J. Digiovanni (2005). "Chemopreventive and therapeutic efficacy of orally active tyrosine kinase inhibitors in a transgenic mouse model of gallbladder carcinoma." Clin Cancer Res **11**(15): 5572-5580.

Kim, H. J., M. H. Kim, S. J. Myung, B. C. Lim, E. T. Park, K. S. Yoo, D. W. Seo, S. K. Lee and Y. I. Min (1999). "A new strategy for the application of CA19-9 in the differentiation of pancreaticobiliary cancer: analysis

using a receiver operating characteristic curve." Am J Gastroenterol **94**(7): 1941-1946.

Klump, B., C. J. Hsieh, S. Dette, K. Holzmann, R. Kiebetalich, M. Jung, U. Sinn, M. Ortner, R. Porschen and M. Gregor (2003). "Promoter methylation of INK4a/ARF as detected in bile-significance for the differential diagnosis in biliary disease." Clin Cancer Res **9**(5): 1773-1778.

Ko, A. H., J. Scott, M. A. Tempero and J. W. PArk (2007). "Detection and significance of circulating tumor cells (CTC) in patients with metastatic pancreatic cancer (PC) receiving systemic therapy. ." J Clin Oncol **25**(18 (Supp, ASCO abstracts)).

Koprowski, H., Z. Steplewski, K. Mitchell, M. Herlyn, D. Herlyn and P. Fuhrer (1979). "Colorectal carcinoma antigens detected by hybridoma antibodies." Somatic Cell Genet **5**(6): 957-971.

Kristiansen, T. Z., H. C. Harsha, M. Gronborg, A. Maitra and A. Pandey (2008). "Differential membrane proteomics using 18O-labeling to identify biomarkers for cholangiocarcinoma." J Proteome Res **7**(11): 4670-4677.

Kubicka, S., F. Kuhnel, P. Flemming, B. Hain, N. Kezmic, K. L. Rudolph, M. Manns and P. N. Meier (2001). "K-ras mutations in the bile of patients with primary sclerosing cholangitis." Gut **48**(3): 403-408.

Kubota, Y., M. Takaoka, K. Tani, M. Ogura, H. Kin, K. Fujimura, T. Mizuno and K. Inoue (1993). "Endoscopic transpapillary biopsy for diagnosis of patients with pancreaticobiliary ductal strictures." Am J Gastroenterol **88**(10): 1700-1704.

Kurzawinski, T. R., A. Deery, J. S. Dooley, R. Dick, K. E. Hobbs and B. R. Davidson (1993). "A prospective study of biliary cytology in 100 patients with bile duct strictures." Hepatology **18**(6): 1399-1403.

Lee, H. Y., S. H. Kim, J. M. Lee, S. W. Kim, J. Y. Jang, J. K. Han and B. I. Choi (2006). "Preoperative assessment of resectability of hepatic hilar cholangiocarcinoma: combined CT and cholangiography with revised criteria." Radiology **239**(1): 113-121.

Lee, J. G., J. W. Leung, J. Baillie, L. J. Layfield and P. B. Cotton (1995). "Benign, dysplastic, or malignant--making sense of endoscopic bile duct brush cytology: results in 149 consecutive patients." Am J Gastroenterol **90**(5): 722-726.

Leelawat, K., S. Leelawat, P. Tepaksorn, P. Rattanasinganchan, A. Leungchaweng, R. Tohtong and P. Sobhon (2006). "Involvement of c-Met/hepatocyte growth factor pathway in cholangiocarcinoma cell invasion and its therapeutic inhibition with small interfering RNA specific for c-Met." J Surg Res **136**(1): 78-84.

Lehmann, U. and H. Kreipe (2001). "Real-time PCR analysis of DNA and RNA extracted from formalin-fixed and paraffin-embedded biopsies." Methods **25**(4): 409-418.

Leja, J., A. Essaghir, M. Essand, K. Wester, K. Oberg, T. H. Totterman, R. Lloyd, G. Vasmatzis, J. B. Demoulin and V. Giandomenico (2009). "Novel markers for enterochromaffin cells and gastrointestinal neuroendocrine carcinomas." Mod Pathol **22**(2): 261-272.

Levy, C., J. Lymp, P. Angulo, G. J. Gores, N. Larusso and K. D. Lindor (2005). "The value of serum CA 19-9 in predicting

cholangiocarcinomas in patients with primary sclerosing cholangitis." Dig Dis Sci **50**(9): 1734-1740.

Lewis, J. T., J. A. Talwalkar, C. B. Rosen, T. C. Smyrk and S. C. Abraham (2010). "Precancerous bile duct pathology in end-stage primary sclerosing cholangitis, with and without cholangiocarcinoma." Am J Surg Pathol **34**(1): 27-34.

Lim, J. H. and C. K. Park (2004). "Pathology of cholangiocarcinoma." Abdom Imaging **29**(5): 540-547.

Lindberg, B., U. Arnelo, A. Bergquist, A. Thorne, A. Hjerpe, S. Granqvist, L. O. Hansson, B. Tribukait, B. Persson and U. Broome (2002). "Diagnosis of biliary strictures in conjunction with endoscopic retrograde cholangiopancreaticography, with special reference to patients with primary sclerosing cholangitis." Endoscopy **34**(11): 909-916.

Linton, K. M., Y. Hey, E. Saunders, M. Jeziorska, J. Denton, C. L. Wilson, R. Swindell, S. Dibben, C. J. Miller, S. D. Pepper, J. A. Radford and A. J. Freemont (2008). "Acquisition of biologically relevant gene expression data by Affymetrix microarray analysis of archival formalin-fixed paraffin-embedded tumours." Br J Cancer **98**(8): 1403-1414.

Liu, L., G. Q. Liao, P. He, H. Zhu, P. H. Liu, Y. M. Qu, X. M. Song, Q. W. Xu, Q. Gao, Y. Zhang, W. F. Chen and Y. H. Yin (2008). "Detection of circulating cancer cells in lung cancer patients with a panel of marker genes." Biochem Biophys Res Commun **372**(4): 756-760.

Liu, X. F., H. Zhang, S. G. Zhu, X. T. Zhou, H. L. Su, Z. Xu and S. J. Li (2006). "Correlation of p53 gene mutation and expression of P53 protein in cholangiocarcinoma." World J Gastroenterol **12**(29): 4706-4709.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**(4): 402-408.

Logrono, R. and I. Waxman (2001). "Interactive role of the cytopathologist in EUS-guided fine needle aspiration: an efficient approach." Gastrointest Endosc **54**(4): 485-490.

Macken, E., M. Drijkoningen, E. Van Aken and W. Van Steenbergen (2000). "Brush cytology of ductal strictures during ERCP." Acta Gastroenterol Belg **63**(3): 254-259.

Madonia, S., E. Aragona, S. Maisano, L. Montalbano, M. Olivo, F. Rossi, G. Restivo and M. Cottone (2007). "CA 19-9 to rule out pancreatic or biliary cancer among patients with cholestasis: an unsuitable test?" Dig Dis Sci **52**(4): 1125-1127.

Maeda, T., K. Kajiyama, E. Adachi, K. Takenaka, K. Sugimachi and M. Tsuneyoshi (1996). "The expression of cytokeratins 7, 19, and 20 in primary and metastatic carcinomas of the liver." Mod Pathol **9**(9): 901-909.

Maeda, T., P. Sepe, S. Lahousse, S. Tamaki, M. Enjoji, J. R. Wands and S. M. de la Monte (2003). "Antisense oligodeoxynucleotides directed against aspartyl (asparaginyl) beta-hydroxylase suppress migration of cholangiocarcinoma cells." J Hepatol **38**(5): 615-622.

Magnusson, M., A. C. Brun, N. Miyake, J. Larsson, M. Ehinger, J. M. Bjornsson, A. Wutz, M. Sigvardsson and S. Karlsson (2007). "HOXA10 is a critical regulator for hematopoietic stem cells and erythroid/megakaryocyte development." Blood **109**(9): 3687-3696.

Mansfield, J. C., S. M. Griffin, V. Wadehra and K. Matthewson (1997). "A prospective evaluation of cytology from biliary strictures." Gut **40**(5): 671-677.

Mansfield, S. D., O. Barakat, R. M. Charnley, B. C. Jaques, C. B. O'Suilleabhain, P. J. Atherton and D. Manas (2005). "Management of hilar cholangiocarcinoma in the North of England: pathology, treatment, and outcome." World J Gastroenterol **11**(48): 7625-7630.

Masuda, N., T. Ohnishi, S. Kawamoto, M. Monden and K. Okubo (1999). "Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples." Nucleic Acids Res **27**(22): 4436-4443.

Matull, W. R., F. Andreola, A. Loh, Z. Adiguzel, M. Deheragoda, U. Qureshi, S. K. Batra, D. M. Swallow and S. P. Pereira (2008). "MUC4 and MUC5AC are highly specific tumour-associated mucins in biliary tract cancer." Br J Cancer **98**(10): 1675-1681.

Matull, W. R., S. A. Khan and S. P. Pereira (2007). "Re: Impact of classification of hilar cholangiocarcinomas (Klatskin tumors) on incidence of intra- and extrahepatic cholangiocarcinoma in the United States." J Natl Cancer Inst **99**(5): 407; author reply 407-408.

Meng, F., R. Henson, M. Lang, H. Wehbe, S. Maheshwari, J. T. Mendell, J. Jiang, T. D. Schmittgen and T. Patel (2006). "Involvement of human micro-RNA in growth and response to chemotherapy in human cholangiocarcinoma cell lines." Gastroenterology **130**(7): 2113-2129.

Mengual, L., M. Burset, E. Ars, M. J. Ribal, J. J. Lozano, B. Minana, L. Sumoy and A. Alcaraz (2006). "Partially degraded RNA from bladder washing

is a suitable sample for studying gene expression profiles in bladder cancer." Eur Urol **50**(6): 1347-1355; discussion 1355-1346.

Menzel, J., C. Poremba, K. H. Dietl and W. Domschke (2000). "Preoperative diagnosis of bile duct strictures--comparison of intraductal ultrasonography with conventional endosonography." Scand J Gastroenterol **35**(1): 77-82.

Mills, R., M. H. Chapman, N. S. Sandanayake and S. P. Pereira (2009). High Success Rate for Tissue Diagnosis in 198 patients with Biliary Tract Cancer. BASL Annual General Meeting. London.

Mills, R., M. H. Chapman, N. S. Sandanayake and S. P. Pereira (2009). High Success Rate for Tissue Diagnosis in 198 patients with Biliary Tract Cancer. BASL Annual General Meeting. London. **Abstract**.

Miralles, C., M. Orea, P. Espana, M. Provencio, A. Sanchez, B. Cantos, R. Cubedo, E. Carcereny, F. Bonilla and T. Gea (2003). "Cancer antigen 125 associated with multiple benign and malignant pathologies." Ann Surg Oncol **10**(2): 150-154.

Mitchell, P. S., R. K. Parkin, E. M. Kroh, B. R. Fritz, S. K. Wyman, E. L. Pogosova-Agadjanyan, A. Peterson, J. Noteboom, K. C. O'Briant, A. Allen, D. W. Lin, N. Urban, C. W. Drescher, B. S. Knudsen, D. L. Stirewalt, R. Gentleman, R. L. Vessella, P. S. Nelson, D. B. Martin and M. Tewari (2008). "Circulating microRNAs as stable blood-based markers for cancer detection." Proc Natl Acad Sci U S A **105**(30): 10513-10518.

Mohandas, K. M., V. S. Swaroop, S. U. Gullar, U. R. Dave, P. Jagannath and L. J. DeSouza (1994). "Diagnosis of malignant obstructive jaundice by

bile cytology: results improved by dilating the bile duct strictures." Gastrointest Endosc **40**(2 Pt 1): 150-154.

Morales, C. P., J. S. Burdick, M. H. Saboorian, W. E. Wright and J. W. Shay (1998). "In situ hybridization for telomerase RNA in routine cytologic brushings for the diagnosis of pancreaticobiliary malignancies." Gastrointest Endosc **48**(4): 402-405.

Moreno Luna, L. E. and G. J. Gores (2006). "Advances in the diagnosis of cholangiocarcinoma in patients with primary sclerosing cholangitis." Liver Transpl **12**(11 Suppl 2): S15-19.

Nagrath, S., L. V. Sequist, S. Maheswaran, D. W. Bell, D. Irimia, L. Ulkus, M. R. Smith, E. L. Kwak, S. Digumarthy, A. Muzikansky, P. Ryan, U. J. Balis, R. G. Tompkins, D. A. Haber and M. Toner (2007). "Isolation of rare circulating tumour cells in cancer patients by microchip technology." Nature **450**(7173): 1235-1239.

Nakagawa, H., S. Liyanarachchi, R. V. Davuluri, H. Auer, E. W. Martin, Jr., A. de la Chapelle and W. L. Frankel (2004). "Role of cancer-associated stromal fibroblasts in metastatic colon cancer to the liver and their expression profiles." Oncogene **23**(44): 7366-7377.

Nakanuma, S., H. Tajima, K. Okamoto, H. Hayashi, H. Nakagawara, I. Onishi, H. Takamura, H. Kitagawa, S. Fushida, T. Tani, T. Fujimura, M. Kayahara, T. Ohta, T. Wakayama, S. Iseki and S. Harada (2010). "Tumor-derived trypsin enhances proliferation of intrahepatic cholangiocarcinoma cells by activating protease-activated receptor-2." Int J Oncol **36**(4): 793-800.

Nakata, B., Y. S. Chung, Y. Kato, M. Ogawa, Y. Ogawa, A. Inui, K. Maeda, T. Sawada and M. Sowa (1996). "Clinical significance of serum CYFRA 21-1 in gastric cancer." Br J Cancer **73**(12): 1529-1532.

Nakata, B., T. Takashima, Y. Ogawa, T. Ishikawa and K. Hirakawa (2004). "Serum CYFRA 21-1 (cytokeratin-19 fragments) is a useful tumour marker for detecting disease relapse and assessing treatment efficacy in breast cancer." Br J Cancer **91**(5): 873-878.

Navin, N., A. Krasnitz, L. Rodgers, K. Cook, J. Meth, J. Kendall, M. Riggs, Y. Eberling, J. Troge, V. Grubor, D. Levy, P. Lundin, S. Maner, A. Zetterberg, J. Hicks and M. Wigler (2010). "Inferring tumor progression from genomic heterogeneity." Genome Res **20**(1): 68-80.

Ng, E. K., W. W. Chong, H. Jin, E. K. Lam, V. Y. Shin, J. Yu, T. C. Poon, S. S. Ng and J. J. Sung (2009). "Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening." Gut **58**(10): 1375-1381.

Nichols, J. C., G. J. Gores, N. F. LaRusso, R. H. Wiesner, D. M. Nagorney and R. E. Ritts, Jr. (1993). "Diagnostic role of serum CA 19-9 for cholangiocarcinoma in patients with primary sclerosing cholangitis." Mayo Clin Proc **68**(9): 874-879.

O'Hara, S. M., J. G. Moreno, D. R. Zweitzig, S. Gross, L. G. Gomella and L. W. Terstappen (2004). "Multigene reverse transcription-PCR profiling of circulating tumor cells in hormone-refractory prostate cancer." Clin Chem **50**(5): 826-835.

Obama, K., K. Ura, M. Li, T. Katagiri, T. Tsunoda, A. Nomura, S. Satoh, Y. Nakamura and Y. Furukawa (2005). "Genome-wide analysis of gene

expression in human intrahepatic cholangiocarcinoma." Hepatology **41**(6): 1339-1348.

Ohlund, D., C. Lundin, B. Ardnor, M. Oman, P. Naredi and M. Sund (2009). "Type IV collagen is a tumour stroma-derived biomarker for pancreas cancer." Br J Cancer **101**(1): 91-97.

Okamura, N., M. Yoshida, A. Shibuya, H. Sugiura, I. Okayasu and M. Ohbu (2005). "Cellular and stromal characteristics in the scirrhous hepatocellular carcinoma: comparison with hepatocellular carcinomas and intrahepatic cholangiocarcinomas." Pathol Int **55**(11): 724-731.

Paik, W. H., Y. S. Park, J. H. Hwang, S. H. Lee, C. J. Yoon, S. G. Kang, J. K. Lee, J. K. Ryu, Y. T. Kim and Y. B. Yoon (2009). "Palliative treatment with self-expandable metallic stents in patients with advanced type III or IV hilar cholangiocarcinoma: a percutaneous versus endoscopic approach." Gastrointest Endosc **69**(1): 55-62.

Paska, C., K. Bogi, L. Szilak, A. Tokes, E. Szabo, I. Sziller, J. Rigo, Jr., G. Sobel, I. Szabo, P. Kaposi-Novak, A. Kiss and Z. Schaff (2004). "Effect of formalin, acetone, and RNAlater fixatives on tissue preservation and different size amplicons by real-time PCR from paraffin-embedded tissue." Diagn Mol Pathol **13**(4): 234-240.

Patel, A. H., D. M. Harnois, G. G. Klee, N. F. LaRusso and G. J. Gores (2000). "The utility of CA 19-9 in the diagnoses of cholangiocarcinoma in patients without primary sclerosing cholangitis." Am J Gastroenterol **95**(1): 204-207.

Penland, S. K., T. O. Keku, C. Torrice, X. He, J. Krishnamurthy, K. A. Hoadley, J. T. Woosley, N. E. Thomas, C. M. Perou, R. S. Sandler and

N. E. Sharpless (2007). "RNA expression analysis of formalin-fixed paraffin-embedded tumors." Lab Invest **87**(4): 383-391.

Perou, C. M., T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown and D. Botstein (2000). "Molecular portraits of human breast tumours." Nature **406**(6797): 747-752.

Ponchon, T., P. Gagnon, F. Berger, M. Labadie, A. Liaras, A. Chavaillon and R. Bory (1995). "Value of endobiliary brush cytology and biopsies for the diagnosis of malignant bile duct stenosis: results of a prospective study." Gastrointest Endosc **42**(6): 565-572.

Ponsioen, C. Y., S. M. Vrouenraets, A. W. van Milligen de Wit, P. Sturm, M. Tascilar, G. J. Offerhaus, M. Prins, K. Huibregtse and G. N. Tytgat (1999). "Value of brush cytology for dominant strictures in primary sclerosing cholangitis." Endoscopy **31**(4): 305-309.

proteinatlas.org "Human-Protein-Atlas." Online database.

Pugliese, V., M. Conio, G. Nicolo, S. Saccomanno and B. Gatteschi (1995). "Endoscopic retrograde forceps biopsy and brush cytology of biliary strictures: a prospective study." Gastrointest Endosc **42**(6): 520-526.

Pujol, J. L., O. Molinier, W. Ebert, J. P. Daures, F. Barlesi, G. Buccheri, M. Paesmans, E. Quoix, D. Moro-Sibilot, M. Szturmowicz, J. M. Brechot, T. Muley and J. Grenier (2004). "CYFRA 21-1 is a prognostic determinant in non-small-cell lung cancer: results of a meta-analysis in 2063 patients." Br J Cancer **90**(11): 2097-2105.

Ramage, J. K., A. Donaghy, J. M. Farrant, R. Iorns and R. Williams (1995). "Serum tumor markers for the diagnosis of cholangiocarcinoma in primary sclerosing cholangitis." Gastroenterology **108**(3): 865-869.

Rea, D. J., J. K. Heimbach, C. B. Rosen, M. G. Haddock, S. R. Alberts, W. K. Kremers, G. J. Gores and D. M. Nagorney (2005). "Liver transplantation with neoadjuvant chemoradiation is more effective than resection for hilar cholangiocarcinoma." Ann Surg **242**(3): 451-458; discussion 458-461.

Rhodes, J. M. (1999). "Usefulness of novel tumour markers." Ann Oncol **10 Suppl 4**: 118-121.

Ribeiro-Silva, A., H. Zhang and S. S. Jeffrey (2007). "RNA extraction from ten year old formalin-fixed paraffin-embedded breast cancer samples: a comparison of column purification and magnetic bead-based technologies." BMC Mol Biol **8**: 118.

Romagnuolo, J., M. Bardou, E. Rahme, L. Joseph, C. Reinhold and A. N. Barkun (2003). "Magnetic resonance cholangiopancreatography: a meta-analysis of test performance in suspected biliary disease." Ann Intern Med **139**(7): 547-557.

Rosch, T., K. Hofrichter, E. Frimberger, A. Meining, P. Born, N. Weigert, H. D. Allescher, M. Classen, M. Barbur, U. Schenck and M. Werner (2004). "ERCP or EUS for tissue diagnosis of biliary strictures? A prospective comparative study." Gastrointest Endosc **60**(3): 390-396.

Rosch, T., A. Meining, S. Fruhmorgen, C. Zillinger, V. Schusdziarra, K. Hellerhoff, M. Classen and H. Helmberger (2002). "A prospective

comparison of the diagnostic accuracy of ERCP, MRCP, CT, and EUS in biliary strictures." Gastrointest Endosc **55**(7): 870-876.

Saijyo, S., T. Kudo, M. Suzuki, Y. Katayose, M. Shinoda, T. Muto, K. Fukuhara, T. Suzuki and S. Matsuno (1995). "Establishment of a new extrahepatic bile duct carcinoma cell line, TFK-1." Tohoku J Exp Med **177**(1): 61-71.

Sarno, J. L., H. J. Kliman and H. S. Taylor (2005). "HOXA10, Pbx2, and Meis1 protein expression in the human endometrium: formation of multimeric complexes on HOXA10 target genes." J Clin Endocrinol Metab **90**(1): 522-528.

Saurin, J. C., M. O. Joly-Pharaboz, P. Pernas, L. Henry, T. Ponchon and J. J. Madjar (2000). "Detection of Ki-ras gene point mutations in bile specimens for the differential diagnosis of malignant and benign biliary strictures." Gut **47**(3): 357-361.

Schoefl, R., M. Haefner, F. Wrba, F. Pfeffel, C. Stain, R. Poetzi and A. Gangl (1997). "Forceps biopsy and brush cytology during endoscopic retrograde cholangiopancreatography for the diagnosis of biliary stenoses." Scand J Gastroenterol **32**(4): 363-368.

Siqueira, E., R. E. Schoen, W. Silverman, J. Martin, M. Rabinovitz, J. L. Weissfeld, K. Abu-Elmaagd, J. R. Madariaga and A. Slivka (2002). "Detecting cholangiocarcinoma in patients with primary sclerosing cholangitis." Gastrointest Endosc **56**(1): 40-47.

Sirica, A. E., M. H. Nathanson, G. J. Gores and N. F. Larusso (2008). "Pathobiology of biliary epithelia and cholangiocarcinoma: proceedings

of the Henry M. and Lillian Stratton Basic Research Single-Topic Conference." Hepatology **48**(6): 2040-2046.

Sitek, B., B. Sipos, I. Alkatout, G. Poschmann, C. Stephan, T. Schulenborg, K. Marcus, J. Luttges, D. D. Dittert, G. Baretton, W. Schmiegel, S. A. Hahn, G. Kloppel, H. E. Meyer and K. Stuhler (2009). "Analysis of the pancreatic tumor progression by a quantitative proteomic approach and immunhistochemical validation." J Proteome Res **8**(4): 1647-1656.

Smirnov, D. A., D. R. Zweitzig, B. W. Foulk, M. C. Miller, G. V. Doyle, K. J. Pienta, N. J. Meropol, L. M. Weiner, S. J. Cohen, J. G. Moreno, M. C. Connelly, L. W. Terstappen and S. M. O'Hara (2005). "Global gene expression profiling of circulating tumor cells." Cancer Res **65**(12): 4993-4997.

Soeth, E., U. Grigoleit, B. Moellmann, C. Roder, B. Schniewind, B. Kremer, H. Kalthoff and I. Vogel (2005). "Detection of tumor cell dissemination in pancreatic ductal carcinoma patients by CK 20 RT-PCR indicates poor survival." J Cancer Res Clin Oncol **131**(10): 669-676.

Spira, A., J. E. Beane, V. Shah, K. Steiling, G. Liu, F. Schembri, S. Gilman, Y. M. Dumas, P. Calner, P. Sebastiani, S. Sridhar, J. Beamis, C. Lamb, T. Anderson, N. Gerry, J. Keane, M. E. Lenburg and J. S. Brody (2007). "Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer." Nat Med **13**(3): 361-366.

Start, R. D., S. S. Cross and J. H. Smith (1992). "Assessment of specimen fixation in a surgical pathology service." J Clin Pathol **45**(6): 546-547.

Steinberg, W. (1990). "The clinical utility of the CA 19-9 tumor-associated antigen." Am J Gastroenterol **85**(4): 350-355.

Stoeber, K., R. Swinn, A. T. Prevost, P. de Clive-Lowe, I. Halsall, S. M. Dilworth, J. Marr, W. H. Turner, N. Bullock, A. Doble, C. N. Hales and G. H. Williams (2002). "Diagnosis of genito-urinary tract cancer by detection of minichromosome maintenance 5 protein in urine sediments." J Natl Cancer Inst **94**(14): 1071-1079.

Sugiyama, M., Y. Atomi, N. Wada, A. Kuroda and T. Muto (1996). "Endoscopic transpapillary bile duct biopsy without sphincterotomy for diagnosing biliary strictures: a prospective comparative study with bile and brush cytology." Am J Gastroenterol **91**(3): 465-467.

Takada, M., N. Masuda, E. Matsuura, Y. Kusunoki, K. Matui, K. Nakagawa, T. Yana, I. Tuyuguchi, I. Oohata and M. Fukuoka (1995). "Measurement of cytokeratin 19 fragments as a marker of lung cancer by CYFRA 21-1 enzyme immunoassay." Br J Cancer **71**(1): 160-165.

Tamada, K., N. Ueno, T. Tomiyama, A. Oohashi, S. Wada, T. Nishizono, S. Tano, T. Aizawa, K. Ido and K. Kimura (1998). "Characterization of biliary strictures using intraductal ultrasonography: comparison with percutaneous cholangioscopic biopsy." Gastrointest Endosc **47**(5): 341-349.

Taniai, M., H. Higuchi, L. J. Burgart and G. J. Gores (2002). "p16INK4a promoter mutations are frequent in primary sclerosing cholangitis (PSC) and PSC-associated cholangiocarcinoma." Gastroenterology **123**(4): 1090-1098.

Taylor-Robinson, S. D., M. B. Toledano, S. Arora, T. J. Keegan, S. Hargreaves, A. Beck, S. A. Khan, P. Elliott and H. C. Thomas (2001).

"Increase in mortality rates from intrahepatic cholangiocarcinoma in England and Wales 1968-1998." Gut **48**(6): 816-820.

Textor, H. J., S. Flacke, D. Pauleit, E. Keller, M. Neubrand, B. Terjung, J. Gieseke, C. Scheurlen, T. Sauerbruch and H. H. Schild (2002). "Three-dimensional magnetic resonance cholangiopancreatography with respiratory triggering in the diagnosis of primary sclerosing cholangitis: comparison with endoscopic retrograde cholangiography." Endoscopy **34**(12): 984-990.

Thongprasert, S. (2005). "The role of chemotherapy in cholangiocarcinoma." Ann Oncol **16 Suppl 2**: ii93-96.

Tidswell, R., M. H. Chapman, N. S. Sandanayake, F. Andreola, G. J. Webster, J. S. Dooley and S. P. Pereira (2009) "Circulating CYFRA 21-1 is a potential marker for cholangiocarcinoma in parimary sclerosing cholangitis." EASL monothematic conference on PSC **Abstract**.

Tillich, M., H. J. Mischinger, K. H. Preisegger, H. Rabl and D. H. Szolar (1998). "Multiphasic helical CT in diagnosis and staging of hilar cholangiocarcinoma." AJR Am J Roentgenol **171**(3): 651-658.

Tischendorf, J. J., M. Kruger, C. Trautwein, N. Duckstein, A. Schneider, M. P. Manns and P. N. Meier (2006). "Cholangioscopic characterization of dominant bile duct stenoses in patients with primary sclerosing cholangitis." Endoscopy **38**(7): 665-669.

Toledano, M. B., A. Mehtan, P. Elliott, S. D. Taylor-Robinson, C. Thomas and S. A. Khan (2009). "Update of the incidence and mortality rates and review of diagnostic trends for intrahepatic cholangiocarcinoma and

selected hepatobiliary tunours in England and Wales." Gut **58**(Suppl): A1-156.

Uenishi, T., O. Yamazaki, H. Tanaka, S. Takemura, T. Yamamoto, S. Tanaka, S. Nishiguchi and S. Kubo (2008). "Serum cytokeratin 19 fragment (CYFRA21-1) as a prognostic factor in intrahepatic cholangiocarcinoma." Ann Surg Oncol **15**(2): 583-589.

Ueno, Y., G. Alpini, K. Yahagi, N. Kanno, Y. Moritoki, K. Fukushima, S. Glaser, G. LeSage and T. Shimosegawa (2003). "Evaluation of differential gene expression by microarray analysis in small and large cholangiocytes isolated from normal mice." Liver Int **23**(6): 449-459.

Utispan, K., P. Thuwajit, Y. Abiko, K. Charngkaew, A. Paupairoj, S. Chau-In and C. Thuwajit (2010). "Gene expression profiling of cholangiocarcinoma-derived fibroblast reveals alterations related to tumor progression and indicates periostin as a poor prognostic marker." Mol Cancer **9**(1): 13.

Valle, J., H. Wasan, D. H. Palmer, D. Cunningham, A. Anthoney, A. Maraveyas, S. Madhusudan, T. Iveson, S. Hughes, S. P. Pereira, M. Roughton and J. Bridgewater (2010). "Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer." N Engl J Med **362**(14): 1273-1281.

Van der Auwera, I., D. Peeters, I. H. Benoy, H. J. Elst, S. J. Van Laere, A. Prove, H. Maes, P. Huget, P. van Dam, P. B. Vermeulen and L. Y. Dirix (2010). "Circulating tumour cell detection: a direct comparison between the CellSearch System, the AdnaTest and CK-19/mammaglobin RT-

PCR in patients with metastatic breast cancer." Br J Cancer **102**(2): 276-284.

Wagner, H. J., K. Knyrim, N. Vakil and K. J. Klose (1993). "Plastic endoprostheses versus metal stents in the palliative treatment of malignant hilar biliary obstruction. A prospective and randomized trial." Endoscopy **25**(3): 213-218.

Wakabayashi, H., S. Akamoto, S. Yachida, K. Okano, K. Izuishi, Y. Nishiyama and H. Maeta (2005). "Significance of fluorodeoxyglucose PET imaging in the diagnosis of malignancies in patients with biliary stricture." Eur J Surg Oncol **31**(10): 1175-1179.

Wang, C., T. Maass, M. Krupp, F. Thieringer, S. Strand, M. A. Worns, A. P. Barreiros, P. R. Galle and A. Teufel (2009). "A systems biology perspective on cholangiocellular carcinoma development: focus on MAPK-signaling and the extracellular environment." J Hepatol **50**(6): 1122-1131.

Wang, Y., Y. Yamaguchi, H. Watanabe, K. Ohtsubo, T. Wakabayashi and N. Sawabu (2002). "Usefulness of p53 gene mutations in the supernatant of bile for diagnosis of biliary tract carcinoma: comparison with K- ras mutation." J Gastroenterol **37**(10): 831-839.

Weber, A., C. von Weyhern, F. Fend, J. Schneider, B. Neu, A. Meining, H. Weidenbach, R. M. Schmid and C. Prinz (2008). "Endoscopic transpapillary brush cytology and forceps biopsy in patients with hilar cholangiocarcinoma." World J Gastroenterol **14**(7): 1097-1101.

Welzel, T. M., K. A. McGlynn, A. W. Hsing, T. R. O'Brien and R. M. Pfeiffer (2006). "Impact of classification of hilar cholangiocarcinomas (Klatskin

tumors) on the incidence of intra- and extrahepatic cholangiocarcinoma in the United States." J Natl Cancer Inst **98**(12): 873-875.

Wise, C., M. Pilanthananond, B. F. Perry, G. Alpini, M. McNeal and S. S. Glaser (2008). "Mechanisms of biliary carcinogenesis and growth." World J Gastroenterol **14**(19): 2986-2989.

Witzigmann, H., F. Berr, U. Ringel, K. Caca, D. Uhlmann, K. Schoppmeyer, A. Tannapfel, C. Wittekind, J. Mossner, J. Hauss and M. Wiedmann (2006). "Surgical and palliative management and outcome in 184 patients with hilar cholangiocarcinoma: palliative photodynamic therapy plus stenting is comparable to r1/r2 resection." Ann Surg **244**(2): 230-239.

Wu, J., F. Lin and Z. Qin (2007). "Sequestration of glyceraldehyde-3-phosphate dehydrogenase to aggregates formed by mutant huntingtin." Acta Biochim Biophys Sin (Shanghai) **39**(11): 885-890.

Wu, Y., F. C. Johlin, S. C. Rayhill, C. S. Jensen, J. Xie, M. B. Cohen and F. A. Mitros (2008). "Long-term, tumor-free survival after radiotherapy combining hepatectomy-Whipple en bloc and orthotopic liver transplantation for early-stage hilar cholangiocarcinoma." Liver Transpl **14**(3): 279-286.

Yamagiwa, Y., F. Meng and T. Patel (2006). "Interleukin-6 decreases senescence and increases telomerase activity in malignant human cholangiocytes." Life Sci **78**(21): 2494-2502.

Yamatoji, M., A. Kasamatsu, Y. Yamano, K. Sakuma, K. Ogoshi, M. Iyoda, K. Shinozuka, K. Ogawara, Y. Takiguchi, M. Shiiba, H. Tanzawa and K. Uzawa (2010). "State of homeobox A10 expression as a putative

prognostic marker for oral squamous cell carcinoma." Oncol Rep **23**(1): 61-67.

Zandrino, F., P. Curone, L. Benzi, M. L. Ferretti and F. Musante (2005). "MR versus multislice CT cholangiography in evaluating patients with obstruction of the biliary tract." Abdom Imaging **30**(1): 77-85.